Dual Degree Dissertation

Sleep-Wake Up Mechanisms for Cellular Heterogeneous Networks

A thesis submitted in partial fulfillment of
the requirements for the degrees of

Bachelor of Technology
in Electrical Engineering
&
Master of Technology
in Communications and Signal Processing

by

DEEKSHA SINHA
(Roll No. 09D07009)

Under the guidance of

Prof. Abhay Karandikar and Prof. V. Kavitha

Department of Electrical Engineering
INDIAN INSTITUTE OF TECHNOLOGY BOMBAY
June, 2014

Dissertation Approval

The dissertation titled

"SLEEP-WAKE UP MECHANISMS
FOR CELLULAR HETEROGENEOUS NETWORKS"

by

Deeksha Sinha
(Roll No. 09D07009)

is approved for the degrees of

Bachelor of Technology in *Electrical Engineering* and
Master of Technology in *Communications and Signal Processing.*

Prof. Prasanna Chaporkar
(Examiner)

Prof. Gaurav Kasbekar
(Examiner)

Prof. Abhay Karandikar
(Guide)

Prof. V. Kavitha
(Co-Guide)

Prof. Bhaskaran Raman
(Chairman)

Date: June 17, 2014

Place: Indian Institute of Technology Bombay, Mumbai

Declaration

I declare that this written submission represents my ideas in my own words and wherever others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/ data/ fact/ source in my submission. I understand that any violation of the above will be cause for disciplinary action by IIT Bombay and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Deeksha

Deeksha Sinha

Date: June 17, 2014

Place: Indian Institute of Technology Bombay, Mumbai

To my parents and teachers

Acknowledgements

I would like to express sincere gratitude towards my supervisors - Prof. V. Kavitha and Prof. Abhay Karandikar for their invaluable guidance throughout the project. Prof. Kavitha's enthusiasm has motivated me immensely to work harder. Her interest in the smallest details of the proofs have been invaluable. She has helped me improve not only the quality of the work but also my style of technical writing.

Prof. Karandikar has helped in emphasizing and understanding the practical significance of the work. His support has been a motivation for me to strive for better results. He has always encouraged and inspired me to achieve more.

I would like to thank my family for their constant support in all my endeavours. Their love has made many difficulties seem easy.

This project has been an amazing journey and undoubtedly one of the best experiences in my stay in IIT Bombay.

Deeksha

June 17, 2014

Deeksha Sinha

Abstract

The Third Generation Partnership Project (3GPP) has proposed the concept of heterogeneous cellular networks to meet the increasing demands of cellular users. These networks consist of small cells like micro, pico and femto cell in addition to the already existing macro cell. The small cells achieve an increase in system capacity by installation of base stations close to the user location. Another major benefit that small cells are expected to bring is improvement in energy efficiency. Proximity of the base station with the user equipment implies lesser power requirement for transmitting information between the two. However, installing a large number of new base stations for the small cells can also lead to an increase in the total energy consumption of the system. Thus, novel approaches for saving energy are required in heterogeneous networks. One of the mechanisms that have been proposed towards this is switching OFF these base stations at times of low load. These mechanisms make use of the variability in the total user demand at different times of the day.

In this work, we consider a heterogeneous network with a macrocell overlaying an area in which pico base stations have also been installed. At times of low load, the pico base stations can be switched OFF to save energy. We study the optimal switch OFF policies for pico base stations. For any given traffic conditions, we determine the maximum fraction of base stations that can be switched OFF while maintaining quality of service (in terms of the average waiting time). Using tools of Multimodularity, we determine the optimal ON-OFF policy for each base station and the optimal association policy between users and base stations.

In the second part of the work, we consider a stand-alone pico base station and focus on its ON-OFF switching depending on the number of users in the coverage region. We find the optimal policy which minimizes a sum of two costs - cost due to power consumption of base station and cost associated with user's Quality of Service.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

The number of mobile subscriptions in the world is increasing at an exponential rate. From 6 billion at the end of 2011, the number is expected to go to 7 billion at the end of 2014 (which is almost 96% of the world's population) [1]. The massive increase in popularity of smart phones and tablets have amounted to huge data traffic in cellular networks. Mobile broadband is expected to be the fastest growing market segment in 2014 [1]. Today, over 79% of the total Internet users access it from their mobile phones [2].

This increase in the number of mobile subscriptions and in the demand of data traffic has led to an enormous increase in the required capacity of cellular systems. New technologies are being proposed to meet this demand. One of the prominent ones is the use of small cells. It has been known that smaller cell sizes typically lead to better system capacities as compared to larger cells. Small cells like femto cells are also ideal for providing high data rates in concentrated regions. Accordingly, the Third Generation Partnership Project (3GPP) standards have introduced the possibility of heterogeneous cellular networks [3].

These increased demands have also led to increased consumption of resources especially the energy resources. Information and Communication Technology (ICT) sector is currently responsible for 2% of the world's total carbon emissions [4]. Even today, a significant portion of the base stations in developing nations are powered by diesel generators which leads to further depletion of already scarce fossil fuels. Further, the energy costs amount to as much as half of the mobile operator's total expenses. Thus, improving energy efficiency of cellular networks has become important both from the operator's profitability point of view as well as to mitigate environmental effects that are being caused. Hence, the current situation requires us to deploy systems with increased capacities and higher energy efficiency.

The energy efficiency of cellular networks can be improved at various levels [4]. Improvements in efficiency of hardware components like power amplifiers lead to lesser power dissipation. Using new technologies like cognitive radio and cooperative relays are also expected to be a step towards green communication. Using sleep-wake up mechanism for base stations reduces power consumption of the network at times of low traffic. We focus on this energy-saving mechanism in this thesis.

1.2 Sleep-Wake Up Mechanisms

A breakdown of the power consumption of a wireless network indicates that base stations consume about 80% of the total energy [5]. Though the energy consumption of the base station varies with the load, there is a major portion which does not depend on the load. Notably, a fixed proportion of energy is consumed even when the base station is serving very few or no users leading to an inefficient system. Thus, there is a scope for making significant energy savings by switching OFF some base stations at the times of low traffic. Numerous sleep-wake up mechanisms [5, 6, 7, 8, 9, 10, 11, 12] have been proposed in the recent years.

These mechanisms take advantage of the conservative deployment of the base stations because base station deployment is typically done to be able to cater to the peak load. Thus, there are large periods of time having less load and which provide scope for switching OFF some base stations.

Sleep-wake up mechanisms become more crucial in heterogeneous networks. In these networks, in addition to the already existing macrocells, there are micro, pico or femto cells as well. As the number of users in any area varies significantly during the day [10], these small cells become redundant at some times. Since the same area is covered by possibly two types of base stations (e.g., macro and a femto base station), one can use both the base stations during peak loads while one (mostly macro base station) is sufficient for low load periods. Thus, the sleep wake mechanism becomes important in the context of small cell/heterogeneous networks.

1.2.1 Review of Proposed Sleep-Wake Up Mechanisms

In the recent years, a number of sleep-wake up mechanisms have been proposed. Different works focus on different aspects of the mechanism - coverage issues when base stations are switched OFF, understanding trade-offs between energy and QoS parameters, compatibility of the mechanism with the current infrastructure and cellular network protocols, etc.

[10] proposes a virtual grid based algorithm. Here, cells which can substitute each other (in terms of the coverage area) are put together in one virtual grid. In each grid, only those many base stations are switched ON as are required for serving the instantaneous traffic. An estimate of the expected traffic in any hour is made using information about the traffic in the previous hour and the traffic at the same time in the previous day. The authors also propose a mechanism for smooth transition between the base stations which are ON in the idle hour and the ones which are gradually turned ON as we approach the peak hours. This algorithm deviates from the optimal algorithm as it ignores the possibility of coordination between base stations belonging to different grids.

In [11], the authors propose a greedy algorithm for switching OFF base stations whenever the traffic load goes below 10% of the peak load. The nearest neighbours of an active base station are sequentially switched OFF while ensuring that the coverage is maintained in 95% of the original coverage area.

[6] proposes a variant of the sleep-wake up concept - cell zooming. In this, cells vary their sizes for achieving load balancing and energy efficiency. If a cell has light load, then it will zoom in and its neighboring cells will simultaneously zoom out to ensure coverage. But the algorithm does not specify the exact condition when a cell should zoom in, it only emphasizes on when the cell can be put in sleep mode. Also, it is unclear if only zooming in and out of cells (without putting them in sleep mode) would make the system more energy efficient.

In [7], the authors suggest switching OFF base stations for which the average traffic is less than a predefined threshold. But this does not incorporate any co-operation among the base stations to share their load. Each base station is evaluated on the basis of total traffic and not just the traffic which cannot be covered by other base stations.

All the above algorithms focus on a homogeneous network setting i.e. a network consisting of only macro-cells. The system dynamics would change due to the introduction of small cells. Some algorithms which are applicable to a heterogeneous setting are described below.

[12] studies the effect of putting base stations in sleep mode on the blocking rate and Quality of Service (QoS) when both 2G and 3G users are present. The authors recommend usage of sleep mode only during the low traffic hours to keep the degradation of QoS to a minimum.

In [8], the authors study wake up mechanisms keeping in mind the features compatible with the hardware of a small cell. Three approaches have been proposed for waking up a small cell in sleep mode - controlled by (1) small cell, (2) core network and (3) user equipment. By default, the small cells are put into sleep mode and are alerted as per the above algorithms. However, the authors do not emphasize on if the small cell should be switched ON whenever there is any user in the vicinity or only if there are users whose demands cannot be met by other base stations which are ON. This can make a significant difference in the energy savings obtained.

In [5], the problem of switching ON and OFF femto base stations is formulated within Markov Decision Process (MDP) framework. The optimal policy that maximises a function of the QoS minus a function of the energy consumed is formulated as a solution to the MDP and can be evaluated using numerical methods. Here, the situations with complete, partial and delayed traffic information are considered separately. In [13], the authors also consider other practical issues like activation time and ping-pong effect which are encountered while deploying sleep-wake up mechanisms. Though the above schemes provide a general solution applicable in many situations, obtaining the MDP solution might not be feasible in real time due to its complexity.

In [9], a system having linearly placed base stations with unidirectional antenna has been considered. For a fixed fraction (η) of base stations to be switched OFF or for each base station to be switched OFF for atleast a given fraction of time (η), the optimal ON-OFF base station policy has been derived. These correspond to two different types of control - centralized and decentralized. Using tools of Multimodularity, it has been shown that the optimal policy in both the cases is a bracket sequence which depends only on the conservation factor (η). This bracket sequence has a very simple form and can be computed easily.

We consider a similar linear deployment of base stations but in a heterogeneous setting containing picocells and a macrocell, with bidirectional antenna and obtain the optimal ON-OFF policy for a given switch-OFF ratio. Further, presence of bidirectional antenna in the pico cells and a macro cell gives users the opportunity to choose among various active base stations. Thus, we focus on the policy that each user should use to choose which base station to connect to.

More importantly we study the finer structural properties of the optimal (bracket) policy and using this study we obtain the optimal fraction to be switched OFF while maintaining QoS given the load conditions. Using the structural properties so obtained, *we have a closed form expression for the average waiting time with the bracket (optimal) ON-OFF policy. In totality, given the load conditions and the QoS constraints, we obtain the optimal operational policies.*

1.3 Objective

Any cellular network aims at providing good quality of service (QoS) to the users. For ensuring good service, for example, it might want to maintain the average waiting time of customers or the call blocking probability below a certain limit. Typically the QoS measures depend both upon the load factor (which depends upon the arrival rate and average work size) of the users as well as the number of available base stations to serve them. As the network also aims to minimize the energy consumption, depending on the traffic, we can optimize on the number of base stations to be operated which would meet the QoS requirements and consume only the minimum required amount of energy.

In the first model, we have considered an area covered by a macro base station, with a major street, carrying significant traffic. Pico base stations (PBSs) have been installed along this street to meet the demands of the users. The aim of the work is to find the maximum fraction of PBSs that can be switched OFF while maintaining the QoS. We also aim to understand which base stations should be switched OFF and among the multiple PBSs and the MBS, which base station should a user connect to. For this a) we first consider the design of optimal ON-OFF and user-base station association policies for the series of PBSs in such a heterogeneous network, given that a fraction η of them need to be switched OFF; b) and then, we evaluate the optimal η to meet the QoS requirements. Towards this, we compute an expression for the desired QoS parameter both in terms of the load factor and the number of active servers (base stations), using queuing theoretic analysis.

In the second model, we consider a single PBS. We aim to find the switch ON-OFF policy of this PBS to minimize the cost of power consumption and a cost related to user inconvenience.

1.4 Organization

The rest of this thesis is organized as follows. In Chapters 2 and 3, we consider the linear deployment scenario of pico base stations and derive the optimal policies (optimal activation policy and optimal user-base station association policy in Chapter 2, optimal switch OFF ratio in Chapter 3). In Chapter 4, we consider decentralized switching policies and derive the optimal activation policy for a pico base station. Proofs of all theorems and lemmas have been provided at the end of the respective chapters.

Chapter 2

Optimal Policies with given Switch OFF Ratio

2.1 System Model

When a cellular network has to be designed in a particular area, its major infrastructural layout is often known which can be exploited to achieve a better deployment. Here one such scenario has been considered, where a busy road passes through a macro cell. Heavy traffic is usually generated on such roads, which can burden the macro base station (MBS). This situation is appropriate for deployment of a heterogeneous network so that the load of the MBS can be shared by a series of PBSs placed along the road. A similar situation arises when a metro line passes through a macro cell. Base stations will need to be installed at various intermediate points along the metro line to cater to the demands of the large number of users who are travelling in the metro. (This linear PBS deployment model has also been considered in works like [14], [15] etc.)

The traffic on the road is generally time varying. For example, the road might carry very heavy traffic during the peak hours in morning and evening while the traffic at mid-day would be much lower. At times of low traffic, some of the base stations can be switched OFF to reduce the energy consumption of the network (This is especially beneficial as picocells might be battery operated). Simultaneously, the network needs to maintain its QoS above a certain acceptable level, irrespective of the traffic conditions. For example, it might want to maintain the average waiting time of a customer within a certain limit, or maintain the call blocking probability for impatient customers below

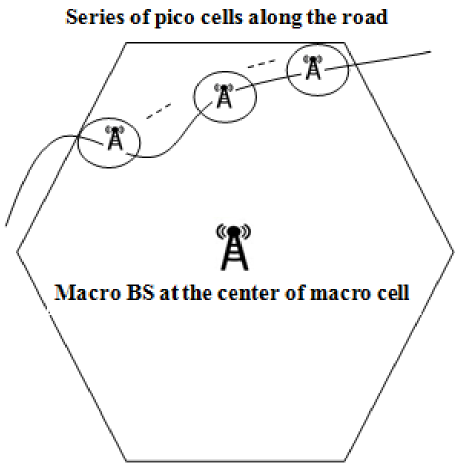


Figure 2.1: System under consideration

a certain limit and so on. The QoS measures depend both upon the arrival rate (load factor) of the users as well the number of available base stations to serve them. We first consider the design of optimal ON-OFF policies for the series of PBSs in such a heterogeneous network, given that a fraction η of these need to be switched OFF. Using queuing theory analysis, the fraction η of the base stations that can be switched OFF at a given time, is obtained by computing an expression for the desired QoS both in terms of the load factor and the number of servers (base stations) and then solving by equating the QoS with the desired/acceptable QoS level.

In our system model, PBSs are placed uniformly (at points $0,\ d,\ 2d, \cdots$) along the street/metro line [1] (which lies in the area covered by the macro cell). Note that the street/line can have curvatures, bends etc as in Figure 2.1. But this can be transformed into a straight line via a homomorphism, as in most of the cases the street is straight locally. Thus, further analysis is done assuming the street to be a straight line.

We assume that users are arriving independently. Also, the rate of arrivals is uniform throughout the line as can be expected on a highway or a metro line.

The analysis here has been done for a constant arrival rate of users and the optimal policies depend on the user arrival rate (λ). Note, however, that for any region, λ varies throughout the day and hence the optimal policies need to be computed once for durations of almost constant λ.

[1] Throughout the below text, 'street' would refer to both metro line and a busy street.

Pico Base Stations

PBSs have been deployed distance d apart from each other. A length of $\frac{d}{2}$ on each side of the PBS is referred to as its cell. Thus, cell of each PBS is of total length d.

Macro Base Station

Pico base stations are required in regions of high load conditions, more so when the MBS is sufficiently far. As the MBS is far from the street, its distance from various points on the street is almost the same. Thus, the rate provided by the MBS to any user depends mainly on the shadowing and fading.

Service Rates

The maximum rate (capacity) at which a BS can serve a user at a distance l (in the absence of any interference) is

$$\hat{\theta}(l) = B \log \left(1 + \frac{P_t}{\sigma^2} \left(1_{\{l \leq l_0\}} + 1_{\{l > l_0\}} \left(\frac{l}{l_0} \right)^{-\beta} \right) \psi |\alpha|^2 \right),$$

where B is the bandwidth, P_t is the transmitted power, σ^2 is the noise variance, β is the pathloss factor, l_0 is the lossless distance, α is a random variable to account for multipath fading and ψ is a random variable to account for shadowing. We assume that there is no inter or intra-cell interference.

Let the set of rates, that the base stations in this network can support, be $\{r_1, r_2, \cdots, r_K\}$. Thus, the rate at which a user at a distance l from the serving BS will be served is

$$\theta = \sum_{i=1}^{K} r_i 1_{\{r_i \leq \hat{\theta}(l) < r_{i+1}\}}.$$

The above gives the rate at which the user is served, which is random because of the fading and shadowing factors.

When served by a PBS, the random rate also depends upon the number of neighbouring PBSs, that are switched OFF as this determines the distance from the serving PBS and hence the transfer rate.

We assume that the shadowing and fading random variables are independent and identically distributed for all users on the street. We define the transfer rate for users in cell n for $n \leq 0$ to be θ^0 where θ^0 is the (random) rate obtained when a user is being served by a PBS in its own cell. Note that θ^0 has same distribution irrespective of the cell number, because of uniform arrivals and identical fading and shadowing distributions in all the cells. That is:

$$\theta^0 = \sum_{i=1}^{K} r_i 1_{\{r_i \leq \hat{\theta}(L^0) < r_{i+1}\}},$$

where L^0 is uniformly distributed between $[0, d/2]$. Similarly, we define θ^k to be the random rate when the serving PBS is k cells away. Here again the distribution depends only upon k and

$$\theta^k = \sum_{i=1}^{K} r_i 1_{\{r_i \leq \hat{\theta}(L^k) < r_{i+1}\}},$$

where L^k is uniformly distributed between $[(2k-1)\frac{d}{2}, (2k+1)\frac{d}{2}]$. Note that one can rewrite $L^k = (2k-1)\frac{d}{2} + 2L^0$. As fading and shadowing remain identical, we conclude that

$$\theta^0 \geq \theta^k \text{ almost surely, } \text{ for all } k \geq 1.$$

Based on similar reasoning, we observe

$$\theta^k \geq \theta^{k+1} \text{ almost surely, } \text{ for all } k \geq 0. \tag{2.1}$$

Similarly, θ^M is defined to be the random rate at which the MBS serves any user on the street. Note that, the distance of different users on the street from the MBS is assumed to be the same. Thus, for different users, the difference in θ^M is only due to different realisations of shadowing and fading.

2.2 Control Policies

We want to determine the maximum throughput that can be achieved by the users when a given fraction of the base station needs to be switched OFF. There are two control policies relevant to this purpose: a) The PBS activation policy which determines which PBSs are OFF b) Given an activation policy, a user-base station association policy which specifies the base station to which a user should connect to.

User-Base Station Association Policy

Each user can connect to any PBS which is ON or the MBS. If the PBS of its own cell is ON, the user obtains service at best rate from its PBS. When this PBS is OFF, an association decision has to be made.

Among all active BSs, a user gets connected to the one from which the received signal strength is the maximum. With high probability, this is the nearest ON PBS or an MBS if all the 'significant' neighbouring PBSs are OFF. Thus we make the following natural choice of parametrized association policies referred to as *J-association policy* : Connect to the MBS if all the J neighbouring PBSs are OFF and if one or more of the J neighbours are ON, then connect to the nearest ON PBS.

To be more precise, we fix a number J and define the association order of the user among the base stations as - PBS at its own position $>$ the two PBSs at distance $d >$ the two PBSs at distance $2d \cdots >$ the two PBSs at distance $Jd >$ macro base station. The user connects to the first active (ON) base station based on this order.

With this association policy, all users in a cell get served by the same base station. Thus, we model the users in a cell as a queue of users getting service from a fixed base station.

PBS can also serve multiple queues of users simultaneously. *We assume that the resources (like channel bandwidth) of the OFF PBSs are appropriately reallocated among the ON PBSs.* Thus, if a PBS is serving multiple queues, its resources increase proportionately so that the rates of the users are not affected by the fact that the PBS is also serving other queues. As the power consumption of the PBS does not vary significantly with load [11], we assume that the power consumption does not change much with this increase in the number of users it is serving.

2.3 System Performance

In this section, we understand the system performance when the activation vector $\mathbf{a}$ (described below) and J-association policy are used. The PBSs are indexed starting from 0. Let the *activation vector* $\mathbf{a} \in \{0,1\}^N$ represent the status of the base stations - $a_i = 1$ if the ith PBS is OFF and it is 0 if the PBS is ON. *We assume that the 0th PBS is always ON.* With large N [2] , this restriction does not alter the performance. Here, bold letters like $\mathbf{a}$ represent an N length sequence while a partial sequence is defined by $\mathbf{a}_j^k := [a_j, a_{j+1}, \cdots, a_k]$.

[2]The major streets or metro lines run over kilometers and the pico cells are usually separated by few hundreds of meters and hence N can be large.

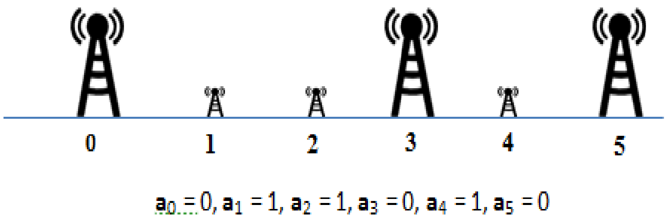


Figure 2.2: Example of an Activation Vector

Performance at n-th point

The users in the n-th cell are served by the PBS s_n cells away where

$$s_n = \inf_{k:|n-k|\leq J} \{|n-k| : a_k = 0\}.$$

If $a_k = 1 \; \forall \; k : |n-k| \leq J$, then they are served by the MBS and $s_n := M$.

Note that the system performance can be controlled only via service rates offered to users in cell n for $n \geq 1$ (as the system consists of PBSs starting from location 0 and the PBS at 0 is always ON).

The queue of users in each cell among $(0, d, 2d, \cdots Nd)$ is served by a BS (PBS or MBS depending upon $(\mathbf{a}, J)$) independent of users in other cells. Thus, we can model each cell as containing an independent queue with Poisson arrivals at rate λ. We *assume that the amount of information S, each user has to transmit is exponentially distributed with mean s.* Given the activation vector $\mathbf{a}$ and J-association policy, a user who arrives in the n-th cell, is served at (random) rate $\theta_n(\mathbf{a}, J) = \theta^{s_n}$. Thus, the user occupies the server of the serving BS for a random time, $S/\theta_n(\mathbf{a}, J)$ and hence we have an M/G/1 queue at n-th point with equivalent service times $B = S/\theta_n(\mathbf{a}, J)$.

Let $w(i)$ represent the expected waiting time of a user in an M/G/1 queue with arrival rate λ and service time $B^i = \frac{S}{\theta^i}$ (recall θ^i is the (random) rate at which a user receives service whose serving PBS is i cells away from it). Note that i can take any value in $[0, \cdots, J]$ and for notational simplicity we also allow $i = M$ to indicate that MBS serves the user. When the queue is not stable then the expected waiting time approaches infinity which we approximate by a large constant H. That is ([16]),

$$w(i) := \begin{cases} \dfrac{\lambda E[(B^i)^2]}{2\,(1-\lambda E[B^i])} & \text{if } \dfrac{1}{E[B^i]} > \lambda \\ H & \text{otherwise.} \end{cases}$$

Thus, in cell n we have an M/G/1 queue with expected waiting time of a user given by :

$$W_n(\mathbf{a}, J) \; = w(s_n(\mathbf{a}, J)). \tag{2.2}$$

By independence, we have

$$E[B^i] = E[S]E\left[\frac{1}{\theta^i}\right] = E[S]\sum_{k=1}^{K}\frac{1}{r_k}1_{\{r_k \le \hat{\theta}(L^i) < r_{k+1}\}}.$$

Using this and (2.1), we get (for $i, i-1 \neq M$)

$$E[B^i] > E[B^{i-1}] \text{ and } E[(B^i)^2] > E[(B^{i-1})^2].$$

Hence, we have $w(i) > w(i-1)$.

Expected Waiting Time of a Typical User

The expected waiting time of a typical user is obtained by first conditioning on its position of arrival and then taking average over the arrival positions. As expected waiting time of all users in a queue is the same, conditioning over arrival position is same as conditioning over the cell in which the user has arrived. By our assumption, a user is equally likely to arrive in any cell. Thus, the expected waiting time of a typical user equals :

$$\frac{1}{N}\sum_{n=0}^{N} W_n(\mathbf{a}, J).$$

We are considering a long street and *hence assume that N is sufficiently large.* Thus, the expected waiting time of a typical user is well approximated by the limit[3]:

$$\overline{W}(\mathbf{a}, J) = \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} W_n(\mathbf{a}, J).$$

(Note $W_0(.,.) = w(0)$ is fixed.)

The Optimization Problem

A network is usually designed to maintain certain desired QoS level throughout the day, irrespective of the time varying load conditions. We take the average of the individual

[3] As the limit may not exist for every activation vector $\mathbf{a}$, we take an upper bound given by the limit superior. We will show that the optimal sequence is periodic, which in turn makes the optimal $\{W_n\}$ periodic and then the limit superior equals the limit.

expected waiting times over all the users (referred to as the customer average) as the QoS metric which needs to be maintained below $\overline{W}_{QoS}$. *The analysis would go through for any other QoS metric that satisfies certain monotonicity properties discussed in the remark of the previous section.*

For any load, there exists a range of switch OFF ratios (η), which meets this QoS requirement. On the other hand, higher η implies higher fraction of PBSs which are switched OFF and thus, lower energy consumption. Thus, we aim to find the maximum fraction of PBSs that can be switched OFF while being able to meet the QoS constraint i.e. the following optimization problem -

$$\sup_{\mathbf{a},J} \ \eta \quad \text{subject to} \tag{2.3}$$
$$\overline{W}(\mathbf{a}, J) \leq \overline{W}_{\text{QoS}} \quad \text{and} \qquad \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} a_n \geq \eta.$$

We solve the above problem in the following manner -

1) For a given switch OFF ratio η, determine the pair $(\mathbf{a}^*(\eta), J^*(\eta))$[4] which minimizes $\overline{W}(\mathbf{a}, J)$ i.e.

$$\min_{\mathbf{a},J} \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} W_n(\mathbf{a}, J)$$
$$\text{subject to} \ \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} a_n \geq \eta. \tag{2.4}$$

2) Obtain the closed form expression for the minimum average waiting time for every η i.e., $\overline{W}(\mathbf{a}^*(\eta), J^*(\eta))$.

3) Determine the maximum η' satisfying the QoS, i.e., $\overline{W}(\mathbf{a}^*(\eta'), J^*(\eta')) \leq \overline{W}_{QoS}$ and show that η' solves (2.3).

2.4 Optimal Activation Vector

As outlined in the previous section, we first consider a fixed switch OFF ratio η and find the activation vector and J-association policy which minimize the customer average of the expected waiting time.

Let $\bar{J}$ represent the distance of the farthest PBS from which the expected waiting

[4] With $N \to \infty$, $\mathbf{a}$ now is in $\{0, 1\}^{\infty}$.

time is better than that provided by the MBS, i.e.

$$\bar{J} := \max_{k} \{k : w(k) > w(M)\}.$$

We begin with the derivation of the optimal activation policy given the J-association policy and the condition that at least η fraction of the PBSs needs to be switched OFF. Towards this, we first consider the following optimization for any fixed $J \leq \bar{J}$:

$$\begin{aligned} &\min_{\mathbf{a}} \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} W_n(\mathbf{a}, J) \\ &\text{subject to} \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} a_n \geq \eta, \end{aligned} \tag{2.5}$$

and show that the optimizer satisfies the constraint with equality.

The solution to this problem is obtained, using concepts of multimodularity. A brief overview of multimodular functions has been presented in Appendix A.

Theorem 1. *For $J \leq \bar{J}$, the function $f_n(\boldsymbol{a}_1^n) := W_{n-J}(\boldsymbol{a}, J)$ is multimodular for every n.* □

Corollary. Any function whose value monotonically increases with the distance between the user and its serving PBS can be shown to be multimodular. Thus, the analysis in this work will also hold for other QoS measures which satisfy this monotonicity condition.

Theorem 2. *For $J \leq \bar{J}$, the solution of (2.5) is given by a bracket sequence $\mathbf{a}^*(\eta)$ for some $\beta \in [0, 1)$ where*

$$\boldsymbol{a}^*(\eta) := \{a_n\}_{n\geq 1}\,;\ a_n = \lfloor n\eta + \beta \rfloor - \lfloor (n-1)\eta + \beta \rfloor.$$ □

In this text, $\lfloor . \rfloor$ and $\lceil . \rceil$ represent the floor and ceil functions respectively.

For the bracket sequence $\mathbf{a}^*(\eta)$, every β defines an optimal policy. Without loss of generality, we consider the policy with $\beta = 0$ for all further discussions.

The bracket sequence $\mathbf{a}^*(\eta)$ in fact, satisfies (2.5) with an equality (Lemma 5.1 in [17]). Thus, the bracket policy is also optimal if the inequality in (2.5) is replaced by an equality.

It is easy to see that if η is rational ($\eta = k_1/k_2$), then the bracket sequence is periodic with period k_2. Since rational numbers are dense in $\mathbb{R}$, we indeed assume a rational η in all our discussions below. Thus, when the optimal policy is used, the ON-OFF pattern of the PBSs will be periodic.

Table 2.1: Example of bracket sequences

η	Periodic sequence in $\mathbf{a}^*(\eta)$
0.2	0 0 0 0 1
0.3	0 0 0 1 0 0 1 0 0 1
0.6	0 1 0 1 1
0.7	0 1 1 0 1 1 0 1 1 1

Further, it should be noted that this optimal bracket policy depends only upon η and is not influenced by other parameters, like J, fading and shadowing distributions etc. This policy also has a very simple and regular form which permits easy calculation.

2.5 Optimal User-Base Station Association Policy

We expect that in the optimal case, the user should connect to the PBS or MBS depending on whichever is able to serve it with less waiting time. As seen in the previous section for any $J \leq \bar{J}$, once the switch OFF ratio η is fixed, the optimal activation policy is independent of J. We now obtain the optimal J with the following theorem.

Theorem 3. $(\boldsymbol{a}^*(\eta), \bar{J})$ *is the minimizer of the optimization problem (2.4) i.e.* $\overline{W}(\boldsymbol{a}, J) \geq \overline{W}(\boldsymbol{a}^*(\eta), \bar{J}) \;\; \forall \; J$ *and for all activation vectors* $\boldsymbol{a}$ *in which the fraction of base stations that are switched OFF is* η. $\square$

Thus given η, the fraction of PBSs to be switched OFF and the policies minimizing the average waiting time are $\mathbf{a}^*(\eta)$ and $\bar{J}$ respectively. Regarding the optimal J, via the above theorem we prove our intuitions correct: for any η it is optimal to connect to the PBSs as long as they provide better service, in terms of the expected waiting time than the MBS (with high probability).

In the next chapter, we find the maximum η *(among rational numbers)* which satisfies the *QoS* requirement *for any given traffic/load conditions (λ and s). The load conditions are reflected via the term $w(i)$ of (2.2).*

2.6 Proofs of Theorems and Lemmas

Proof of Theorem 1. All the sequences in this proof are n length vectors and also J is fixed. Hence, we use the shorthand notation $\mathbf{a}$ in place of $\mathbf{a}_1^n$ for all the vectors and $\theta_n(\mathbf{a})$ in place of $\theta_n(\mathbf{a}, J)$. We need to show

$$f_n(\mathbf{a}+\mathbf{v}) + f_n(\mathbf{a}+\mathbf{u}) \geq f_n(\mathbf{a}) + f_n(\mathbf{a}+\mathbf{u}+\mathbf{v}) \tag{2.6}$$

$\forall\ \mathbf{a} \in \{0,1\}^n$ and $\forall\ \mathbf{u}, \mathbf{v} \in F$ (the multimodular base) with $\mathbf{u} \neq \mathbf{v}$ and such that $\mathbf{a}+\mathbf{u},\ \mathbf{a}+\mathbf{v},\ \mathbf{a}+\mathbf{u}+\mathbf{v} \in \{0,1\}^n$.

Let us define $s_1 = -e_1$ and $s_{n+1} = e_n$. Let $\mathbf{v} = s_j$ and $\mathbf{u} = s_l$ where $j, l \in \{1, n+1\}$ and $l \neq j$. Without loss of generality, we assume that $l > j$.

Consider $j \in \{2, n\}$. Since $\mathbf{a} + \mathbf{v} \in \{0,1\}^n$, we have $a_{j-1} = 0$ and $a_j = 1$. Further,

$$(a+v)_{j-1} = 1, (a+v)_j = 0 \text{ and } a_i = (a+v)_i \quad \forall\ i \neq j, j-1.$$

Therefore, adding $\mathbf{v}$ to $\mathbf{a}$ implies that the PBS in $(j-1)$-th cell which was ON is turned OFF and the PBS in the j-th cell is turned ON.

When $\mathbf{v} = s_1 = -e_1$, we would have

$$a_1 = 1,\ (a+v)_1 = 0 \text{ and } a_i = (a+v)_i \quad \forall\ i \neq 1.$$

Similarly, when $\mathbf{v} = s_{n+1} = e_n$, we would have

$$a_n = 0, (a+v)_n = 1 \text{ and } a_i = (a+v)_i \quad \forall\ i \neq n.$$

Thus, addition of $-e_1$ switches ON the first base station and addition of e_n switches OFF the last base station. All the above will also hold when $\mathbf{v}$ is replaced by $\mathbf{u}$ and j by l. Note that the PBS in the 0th cell is always switched ON. Thus, $a_0 = 0\ \ \forall\ \mathbf{a}$.

Note that l cannot be equal to $j+1$ i.e. $l > j+1$ since $u = s_l$ implies $a_{l-1} = 0$ i.e. $a_j = 0$ if $l = j+1$. But, $v = s_j$ implies $a_j = 1$. Hence, we have an inconsistency if $l = j+1$.

Thus, we only need to consider $j \in [1, n]$, $l \in [2, n+1]\ \forall\ l > j+1$.

Let the closest active (ON) PBS on the left of a user in the$(n-J)$-th cell and the closest active PBS on its right be in cells $K_L(\mathbf{a})$ and $K_R(\mathbf{a})$ respectively i.e.,

$$K_L(\mathbf{a}) = \max_{0 \leq k \leq n-J} \{a_k = 0\} \text{ and}$$

$$K_R(\mathbf{a}) = \begin{cases} n+1 & \text{if } a_i = 1 \ \forall \ n-J \leq i \leq n \\ \min\limits_{k \geq n-J} \{a_k = 0\} & \text{otherwise.} \end{cases}$$

$K_R(\mathbf{a})$ is assigned value $n+1$ when none of the J PBSs to the right of the user are ON. This signifies that the user will be connected either to $K_L(\mathbf{a})$ or the MBS. From now on, we refer to a user in the $(n-J)$-th cell as the $(n-J)$-th user. (All the users in a cell are served by the same BS and have the same expected waiting time, hence the notation.)

Let $B(\mathbf{a})$ represent the base station to which the $(n-J)$-th user is connected. $B(\mathbf{a})$ is either $K_L(\mathbf{a})$ or $K_R(\mathbf{a})$ or the MBS. *Unless there is a change in the base station to which the user is connected, its expected waiting time will not change.*

If $n \leq J$, we will have $n - J \leq 0$. By definition, $\theta_{n-J}(\mathbf{b}) = \theta^0$ for any activation vector $\mathbf{b}$. Thus, almost surely,

$$\theta_{n-J}(\mathbf{a}) = \theta_{n-J}(\mathbf{a}+\mathbf{u}) = \theta_{n-J}(\mathbf{a}+\mathbf{v}) = \theta_{n-J}(\mathbf{a}+\mathbf{u}+\mathbf{v}).$$

$f_n(.)$ represents the expected waiting time of a user in the $(n-J)$-th cell. The rate offered to a user in this cell is θ_{n-J}. If $\theta > \theta'$ almost surely, then $\hat{w}(\theta) > \hat{w}(\theta')$ where $\hat{w}(\theta)$ denotes the expected waiting time of a user who is being served at the random rate θ. Similarly, $\theta = \theta'$ almost surely implies $\hat{w}(\theta) = \hat{w}(\theta')$.[5]

Thus, we have

$$f_n(\mathbf{a}+\mathbf{v}) = f_n(\mathbf{a}+\mathbf{u}) = f_n(\mathbf{a}) = f_n(\mathbf{a}+\mathbf{u}+\mathbf{v}) = w(0),$$

and (2.6) is satisfied. Now, we focus on $n > J$. If

$$j-1 < l-1 < K_L(\mathbf{a}) \text{ or } K_R(\mathbf{a}) < j-1 < l-1 \text{ or}$$

$$j-1 < K_L(\mathbf{a}) \text{ and } (l-1) > K_R(\mathbf{a}),$$

then even after adding $\mathbf{u}$ or $\mathbf{v}$ or $\mathbf{u}$+$\mathbf{v}$ to $\mathbf{a}$, the nearest ON PBS to the $(n-J)$-th user on both its sides remain unchanged. Thus,

$$f_n(\mathbf{a}+\mathbf{v}) = f_n(\mathbf{a}+\mathbf{u}) = f_n(\mathbf{a}) = f_n(\mathbf{a}+\mathbf{u}+\mathbf{v}).$$

Hence, (2.6) is satisfied. Now, let us divide the rest of the possibilities into three scenarios -

[5] We will use this reasoning throughout without mentioning repeatedly.

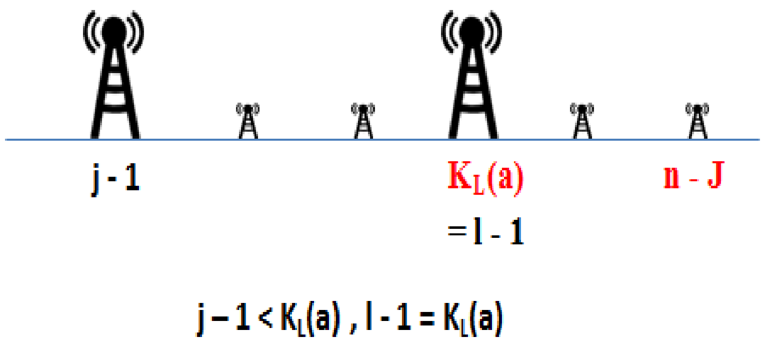
1. $j-1 < K_L(\mathbf{a}),\ l-1 = K_L(\mathbf{a})$ or $l-1 = K_R(\mathbf{a})$

2. $j-1 = K_L(\mathbf{a})$ or $j-1 = K_R(\mathbf{a}),\ l-1 > K_R(\mathbf{a})$

3. $j-1 = K_L(\mathbf{a}),\ l-1 = K_R(\mathbf{a})$.

Considering each of them one-by-one,

Case 1: $j-1 < K_L(\mathbf{a}),\ l-1 = K_L(\mathbf{a})$ or $l-1 = K_R(\mathbf{a})$



As $j-1 < K_L(\mathbf{a})$ switching OFF the PBS in $(j-1)$-th cell and switching ON the PBS in j-th cell or only switching ON the PBS in cell 1 (when $j=1$) will not affect the $(n-J)$-th user. ($j \neq K_L(\mathbf{a})$ as $a_j = 1$ and $a_{K_L(\mathbf{a})} = 0$.)

Thus, $B(\mathbf{a}) = B(\mathbf{a}+\mathbf{v})$.

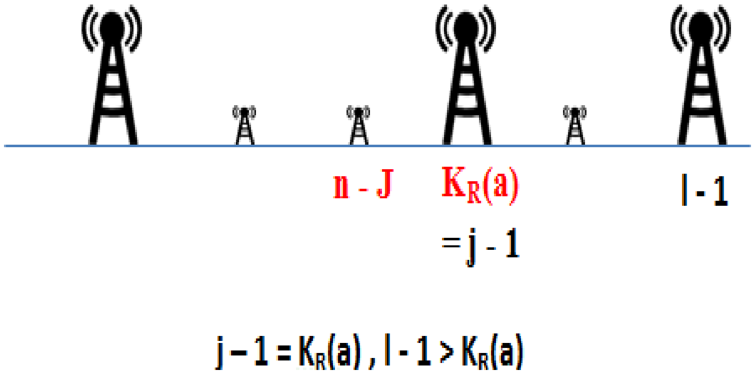
Hence, $f_n(\mathbf{a}+\mathbf{v}) = f_n(\mathbf{a})$ and similarly, $f_n(\mathbf{a}+\mathbf{u}+\mathbf{v}) = f_n(\mathbf{a}+\mathbf{u})$. Therefore, (2.6) is satisfied for all possible values of $\mathbf{v}$ and $\mathbf{u}$ when $j-1 < K_L(\mathbf{a}),\ l-1 = K_L(\mathbf{a})$ or $l-1 = K_R(\mathbf{a})$.

Case 2: $j-1 = K_L(\mathbf{a})$ or $j-1 = K_R(\mathbf{a}), l-1 > K_R(\mathbf{a})$

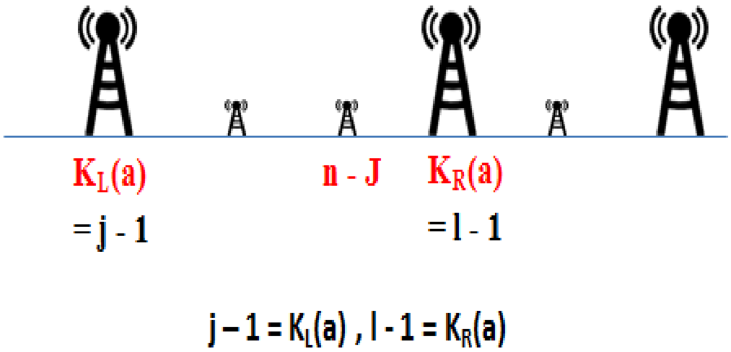
Similar to the previous case, we will now have,

$$f_n(\mathbf{a}+\mathbf{u}) = f_n(\mathbf{a}) \text{ and } f_n(\mathbf{a}+\mathbf{v}+\mathbf{u}) = f_n(\mathbf{a}+\mathbf{v}).$$

Therefore, (2.6) is satisfied for all possible values of $\mathbf{v}$ and $\mathbf{u}$ when $j-1 < K_L(\mathbf{a}),\ l-1 = K_L(\mathbf{a})$.



Case 3: $j - 1 = K_L(\mathbf{a}), l - 1 = K_R(\mathbf{a})$



As $l \leq n + 1$, we have $K_R(\mathbf{a}) \leq n$.

$$\therefore K_R(\mathbf{a}) - (n - J) \leq n - (n - J) \tag{2.7}$$

i.e. the number of cells between the $(n - J)-$th user and $K_R(\mathbf{a})$ is less than or equal to J which implies $(n - J)$-th user is connected to either the PBS at $K_L(\mathbf{a})$ or the PBS at $K_R(\mathbf{a})$. Thus, the user would not be connected to the MBS with this activation vector $\mathbf{a}$. As $K_R(\mathbf{a}) = K_R(\mathbf{a} + \mathbf{v})$, the same arguments hold when the activation vector is $\mathbf{a} + \mathbf{v}$. Thus, even with this activation vector , the user would not be connected to the MBS.

We know that $K_L(\mathbf{a}) \leq n - J$. If $K_L(\mathbf{a}) = n - J$ then by definition, $K_R(\mathbf{a}) = n - J$. Then, $l - 1 = n - J = j - 1$. This is an inconsistency as $l > j + 1$.

Clearly, $K_L(\mathbf{a} + \mathbf{v}) = 1 + K_L(\mathbf{a})$. Thus, when $j - 1 = K_L(\mathbf{a}) < n - J$, addition of $\mathbf{v}$ switches ON a PBS closer to the user. As the expected waiting time monotonically increases with the distance between the user and its serving PBS,

$$f_n(\mathbf{a} + \mathbf{v}) \leq f_n(\mathbf{a}).$$

Using the same arguments we get,

$$f_n(\mathbf{a} + \mathbf{u} + \mathbf{v}) \leq f_n(\mathbf{a} + \mathbf{u}). \tag{2.8}$$

Now, consider the following sub-cases -

1. $K_R(\mathbf{a}) - (n-J) < (n-J) - K_L(\mathbf{a})$

   Here, $B(\mathbf{a}) = K_R(\mathbf{a})$. Also, $K_L(\mathbf{a}+\mathbf{v}) = 1 + K_L(\mathbf{a})$ and $K_R(\mathbf{a}+\mathbf{v}) = K_R(\mathbf{a})$. Thus,

$$K_R(\mathbf{a}+\mathbf{v}) - (n-J) \leq (n-J) - K_L(\mathbf{a}+\mathbf{v}),$$
$$\Rightarrow B(\mathbf{a}+\mathbf{v}) = K_R(\mathbf{a}+\mathbf{v}) = K_R(\mathbf{a}) = B(\mathbf{a}).$$
$$\therefore f_n(\mathbf{a}+\mathbf{v}) = f_n(\mathbf{a}). \tag{2.9}$$

   Adding (2.8) and (2.9), we get

   $f_n(\mathbf{a}+\mathbf{v}) + f_n(\mathbf{a}+\mathbf{u}) \geq f_n(\mathbf{a}) + f_n(\mathbf{a}+\mathbf{u}+\mathbf{v})$.

2. $K_R(\mathbf{a}) - (n-J) \geq (n-J) - K_L(\mathbf{a})$

   Here, $B(\mathbf{a}) = K_L(\mathbf{a})$. Clearly, $K_L(\mathbf{a}) = K_L(\mathbf{a}+\mathbf{u})$ and $K_R(\mathbf{a}+\mathbf{u}) \geq K_R(\mathbf{a})$. So we have,

$$K_R(\mathbf{a}+\mathbf{u}) - (n-J) \geq (n-J) - K_L(\mathbf{a}+\mathbf{u}),$$
$$\Rightarrow B(\mathbf{a}+\mathbf{u}) = K_L(\mathbf{a}+\mathbf{u}) = K_L(\mathbf{a}) = B(\mathbf{a}).$$
$$\therefore f_n(\mathbf{a}+\mathbf{u}) = f_n(\mathbf{a}). \tag{2.10}$$

   Now, let us consider the situation when the activation vector is $\mathbf{a}+\mathbf{v}$. As $K_R(\mathbf{a}) = K_R(\mathbf{a}+\mathbf{v})$, from (2.7), we have

$$K_R(\mathbf{a}+\mathbf{v}) - (n-J) \leq n - (n-J).$$

   Thus, even when the activation vector is $\mathbf{a}+\mathbf{v}$ the user would not be connected to the MBS but would be connected to either $K_R(\mathbf{a}+\mathbf{v})$ or $K_L(\mathbf{a}+\mathbf{v})$.

   Using the hypothesis,

$$K_R(\mathbf{a}+\mathbf{v}) - (n-J) \geq (n-J) - K_L(\mathbf{a}+\mathbf{v}).$$

   Hence, the user is connected to $K_L(\mathbf{a}+\mathbf{v})$. (In the case of equality in the above, user can be connected either to $K_L(\mathbf{a}+\mathbf{v})$ or $K_R(\mathbf{a}+\mathbf{v})$ but this does not make a difference to $f_n(\mathbf{a}+\mathbf{v})$ as both of them are equidistant from the user.) Therefore,

$$f_n(\mathbf{a}+\mathbf{v}+\mathbf{u}) = f_n(\mathbf{a}+\mathbf{v}). \tag{2.11}$$

Adding (2.10) and (2.11), we get

$f_n(\mathbf{a}+\mathbf{v}) + f_n(\mathbf{a}+\mathbf{u}) = f_n(\mathbf{a}) + f_n(\mathbf{a}+\mathbf{u}+\mathbf{v})$.

Thus, for $K_L(\mathbf{a}) = j-1$ and $K_R(\mathbf{a}) = l-1$, (2.6) is satisfied. Therefore, for all $\mathbf{u}, \mathbf{v} \in F$ and $\mathbf{u} \neq \mathbf{v}$, we have proved that

$$f_n(\mathbf{a}+\mathbf{v}) + f_n(\mathbf{a}+\mathbf{u}) \geq f_n(\mathbf{a}) + f_n(\mathbf{a}+\mathbf{u}+\mathbf{v}).$$

Hence, we conclude that $f_n(\mathbf{a}) = W_{n-J}(\mathbf{a})$ is multimodular. □

Proof of Theorem 2. As J remains constant throughout the proof, we will use a shorthand notation of $W_n(\mathbf{a}), \theta_n(\mathbf{a})$ instead of $W_n(\mathbf{a}, J), \theta_n(\mathbf{a}, J)$ respectively. Define

$$f_n(\mathbf{a}_1^n) = W_{n-J}(\mathbf{a}, J).$$

This proof is obtained using Theorem 13. We will verify the validity of its assumptions.

1) $f_n(\mathbf{a}_1^n) = W_{n-J}(\mathbf{a})$ is multimodular from Theorem 1.

2) For assumption 2, we need to prove $\forall\ n > 1$ that

$$f_n(a_1, \cdots, a_n) \geq f_{n-1}(a_2, \cdots, a_n) \text{ i.e.}$$

$$W_{n-J}(a_1, \cdots, a_n, \cdots) \geq W_{(n-1)-J}(a_2, \cdots, a_n, \cdots).$$

If $(n-J) \leq 0$, then the assumption is true because

$$W_{n-J}(a_1, \cdots, a_n, \cdots) = W_{(n-1)-J}(a_2, \cdots, a_n, \cdots) = w(0).$$

Now, let us consider $(n-J) > 0$. Define activation vector

$$\mathbf{c} = (c_1, c_2, \cdots, c_{n-1}, \cdots) := (a_2, \cdots, a_n, \cdots).$$

By our assumption, the element with the index 0 of any activation vector is 0 i.e. $c_0 = a_0 = 0$. Recall that θ_{n-J} is the transfer rate of users in the $(n-J)$-th cell while $\theta_{(n-1)-J}$ is the rate of users in its left neighbouring cell. With the change in activation vector, the status of the PBS originally determined by a_1 is now determined by c_0. Let us call this PBS P. If $a_1 = 0$, then no user other than users of cell 0 would have been connected to PBS in cell 0. As $c_0 = 0$ (by assumption),for all users in cells $1, 2, \cdots$ there will be no change in expected waiting time. If $a_1 = 1$, then P was OFF and now with the new activation vector, it is switched ON. This cannot result in an increase in the expected waiting time of any user in any queue. Therefore, $\forall\ n > 1$

$$W_{n-J}(a_1, \cdots, a_n, \cdots) \geq W_{(n-1)-J}(a_2, \cdots, a_n, \cdots).$$

3) We have $a_0 = 0$. As on adding $\mathbf{b}$, b_{n-m} represents the activation status of the PBS which was originally represented by a_0, we take $\mathbf{b}$ such that $b_{n-m} = 0$. With such a choice clearly,

$$f_n(b_1, \cdots, b_{n-m}, a_1, \cdots, a_m) = f_m(a_1, \cdots, a_m).$$

4) Switching ON a PBS cannot increase a user's expected waiting time i.e.
$W_{n-J}(a_1, \cdots, a_{i-1}, 0, a_{i+1}, \cdots, a_n, \cdot) \leq W_{n-J}(a_1, \cdots, a_{i-1}, 1, a_{i+1}, \cdots, a_n, \cdot) \quad \forall\ i.$
Hence $f_n(\mathbf{a}_1^n) = W_{n-J}(\mathbf{a})$ is increasing in $a_i \quad \forall\ i$. Thus, all the assumptions of Theorem 13 hold . Now,

$$\begin{aligned}
\limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_n(a_1, \cdots, a_n) &= \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} W_{n-J}(\mathbf{a}, J), \\
&= \limsup_{N\to\infty} \frac{1}{N} \left(\sum_{n=1-J}^{0} W_0(\mathbf{a}, J) + \sum_{n=1}^{N-J} W_n(\mathbf{a}, J) \right), \\
&= \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} W_n(\mathbf{a}, J).
\end{aligned}$$

(For $n - J \leq 0, W_{n-J}(\mathbf{a}) = w(0)$.)

Thus, using Theorem 13, the optimization problem (2.5) has the solution as the bracket policy sequence $\mathbf{a}^*(\eta)$. $\square$

Proof of Theorem 3. When the J-association policy is being used, the user being served in the n-th cell can be associated to 0th, 1st, $\cdots$ Jth nearest PBS or the MBS. In the following discussion, by 'user', we refer to the user being served in the nth cell.
Thus, we have the following cases-
1) User is being served by a PBS:
Let the distance of this PBS from the user be l i.e.

$$l(\mathbf{a}, J) := \min_{0 \leq k \leq J} \{a_{n+k} = 0 \text{ or } a_{n-k} = 0\}.$$

Clearly, $l(\mathbf{a}, J) = l(\mathbf{a}, J + 1)$. Therefore,

$$W_n(\mathbf{a}, J + 1) = W_n(\mathbf{a}, J).$$

2) User is being served by the the MBS i.e. $W_n(\mathbf{a}, J) = w(M)$:

$$W_n(\mathbf{a}, J+1) = w(M)\mathbf{1}\left\{a_{n+J+1} = a_{n-J-1} = 1\right\} + w(J+1)\mathbf{1}\left\{a_{n+J+1}a_{n-J-1} = 0\right\},$$

where $\mathbf{1}(.)$ represents the indicator function. Thus,

$$W_n(\mathbf{a}, J+1) \leq W_n(\mathbf{a}, J) \text{ if } J < \bar{J} \text{ and}$$
$$W_n(\mathbf{a}, J+1) \geq W_n(\mathbf{a}, J) \text{ if } J \geq \bar{J}.$$

From the two cases above,

$$W_n(\mathbf{a}, 1) \geq W_n(\mathbf{a}, 2) \geq \cdots W_n(\mathbf{a}, \bar{J}) \text{ and}$$
$$W_n(\mathbf{a}, \bar{J}) \leq W_n(\mathbf{a}, \bar{J}+1) \leq W_n(\mathbf{a}, \bar{J}+2) \leq \cdots .$$

Thus, we have

$$W_n(\mathbf{a}, J) \geq W_n(\mathbf{a}, \bar{J}) \quad \forall\ \mathbf{a},\ J.$$

Averaging over all n's, we get $\overline{W}(\mathbf{a}, J) \geq \overline{W}(\mathbf{a}, \bar{J})\ \forall\ \mathbf{a},\ J$. From Theorem 2, $\overline{W}(\mathbf{a}, \bar{J}) \geq \overline{W}(\mathbf{a}^*, \bar{J})$ for all $\mathbf{a}$ in which the fraction of base stations switched OFF is equal to η. Thus, we have, $\overline{W}(\mathbf{a}, J) \geq \overline{W}(\mathbf{a}^*, \bar{J}) \quad \forall\ J$ and for all $\mathbf{a}$ satisfying the above condition. $\square$

Chapter 3

Optimal Switch OFF Ratio

From the previous chapter, for a given η, the minimum waiting time is given by $\overline{W}(\mathbf{a}^*(\eta), \bar{J})$. Let $\overline{W}^*(\eta) := \overline{W}(\mathbf{a}^*(\eta), \bar{J})$ represent this optimal value for a given η. Note here, $\mathbf{a}^*(\eta)$ depends only upon η. We obtain solution to (2.3) in two steps: a) we obtain an explicit expression for $\overline{W}^*(\eta)$ in terms of η and show that it is monotone in η; b) we then show that the η', that satisfies the equation $\overline{W}^*(\eta') = \overline{W}_{QoS}$, is the required solution.

We study finer structural properties of the bracket policy which helps us to obtain the expression for $\overline{W}^*(\eta)$.

$\overline{W}^*(\eta)$ depends on two factors - the rates at which users are being served in different queues and the frequency of each such rate. As we deal with rational η, we take $\eta = k_1/k_2$ where k_1 and k_2 are integers. We know that the sequence $\mathbf{a}^*(\eta)$ is periodic with period k_2. Thus for any i,

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} W_n(\mathbf{a}^*(\eta), \bar{J}) = \frac{1}{k_2} \sum_{n=i}^{i+k_2-1} W_n(\mathbf{a}^*(\eta), \bar{J}). \tag{3.1}$$

We thus study a block of k_2 consecutive queues. Users in each of these k_2 queues are served by the MBS or by a common PBS which is x cells away from it and $x \in \{0, 1, \cdots \bar{J}\}$.

In the activation vector $\mathbf{a}^*(\eta)$, the fraction of OFF PBSs is exactly equal to η and hence in a block of k_2 PBSs, the number of OFF PBSs will be $\eta \times k_2 = k_1$.

3.1 Analysis of the Bracket Policy

Let us analyze the activation vector $\mathbf{a}^*(\eta)$ which is expressed as $\mathbf{a}^*(\eta) = \{a_n\}_{n\geq 1} = \{\lfloor \eta n \rfloor - \lfloor \eta(n-1) \rfloor\}_{n\geq 1}$. Consider a block of k_2 consecutive PBSs from $n_0 = mk_2$ to

$(m+1)k_2 - 1$ for any integer m. We have,

$$\begin{aligned} a_{n_0} &= \lfloor \eta n_0 \rfloor - \lfloor \eta(n_0 - 1) \rfloor, \\ &= \lfloor mk_1 \rfloor - \lfloor mk_1 - \frac{k_1}{k_2} \rfloor, \\ &= mk_1 - (mk_1 - 1), \\ &= 1. \end{aligned}$$

Thus, the PBS in the n_0 cell is switched OFF. To find the next OFF PBS, we need to find the smallest integer $s > 0$ such that $a_{n_0+s} = 1$ i.e.

$$\lfloor \eta(n_0 + s) \rfloor - \lfloor \eta(n_0 + s - 1) \rfloor = 1.$$

For any integer $s > 0$, we have $\lfloor \eta(n_0 + s - 1) \rfloor \geq mk_1$. Thus, the required s is the smallest integer s with

$$\lfloor \eta(n_0 + s) \rfloor = mk_1 + 1 \text{ and } \lfloor \eta(n_0 + s - 1) \rfloor = mk_1.$$

But

$$\left\lfloor \eta(n_0 + s) \right\rfloor = \left\lfloor \frac{k_1}{k_2}(mk_2 + s) \right\rfloor = mk_1 + \left\lfloor s\frac{k_1}{k_2} \right\rfloor.$$

So, we need the smallest s such that, $\left\lfloor s\frac{k_1}{k_2} \right\rfloor = 1$ and we use the following result:

Lemma 1. $\left\lceil p\frac{k_2}{k_1} \right\rceil$ *is the smallest s such that* $\left\lfloor s\frac{k_1}{k_2} \right\rfloor = p$. □

Thus, after n_0, the next OFF PBS is present in $n_0 + \left\lceil \frac{k_2}{k_1} \right\rceil$ cell.

Proceeding in the same manner, the p-th next OFF PBS can be found by solving for the smallest s such that $\lfloor (mk_2 + s)\frac{k_1}{k_2} \rfloor = mk_1 + p$ and $\lfloor (mk_2 + s - 1)\frac{k_1}{k_2} \rfloor = mk_1 + p - 1$. Using Lemma 1 again, we get $s = \lceil p\frac{k_2}{k_1} \rceil$. Thus,

Lemma 2. *For $n_0 = mk_2$, where m is an integer, $\eta = \frac{k_1}{k_2}$ and $i > 0$,*

$$\begin{aligned} \boldsymbol{a}_{n_0+i} &= 1 \textit{ if } i = \left\lceil \frac{p}{\eta} \right\rceil \textit{ for some } p \in \{1, 2 \cdots\} \\ &= 0 \textit{ otherwise.} \end{aligned}$$ □

We call a queue to be of type j if the nearest ON PBS is j cells away from it.

Let the total number of types of queues in the block be $l(\eta)$. It is easy to see that if there exists a queue of type j, then there will exist queues of types i whenever $0 \leq i \leq j$.

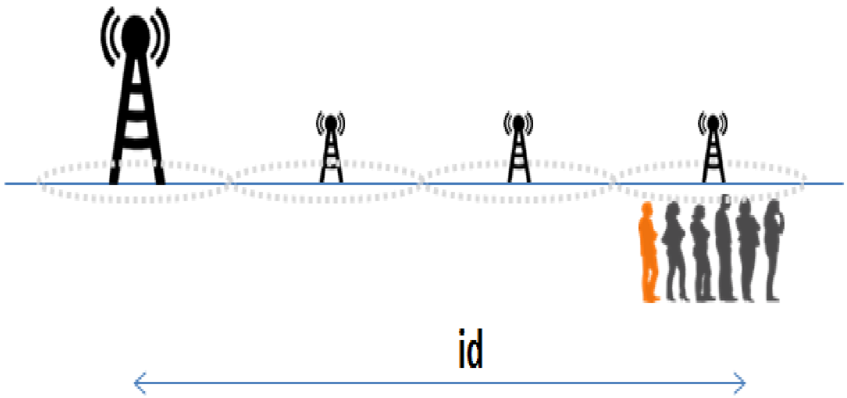


Figure 3.1: Queue of Type i

Thus, $l(\eta) = i$ would mean that the set of possible distances (number of cells) between users and their nearest ON PBS is $\{0, 1, \cdots, i-1\}$ and for every distance in this set, there will exist at least one queue of users at this distance from its nearest serving PBS. We have the following result -

Lemma 3. *$l(\eta) = r+1$ for $h(r-1) < \eta \leq h(r)$ where r is an integer and $h(r) := \dfrac{2r}{1+2r}$.*
□

3.2 Determining Frequency of each Type of Queue

Having found the different types of possible queues for a given η, we need to determine the number of each type of queue in the k_2 block. We obtain these frequencies and then the final expression for $\overline{W}^*(\eta)$ in the following:

Theorem 4. *When $\eta = k_1/k_2$ and $h(r-1) < \eta \leq h(r)$ for some r, with $\gamma := \min\{r-1, \bar{J}\}$ we have the following in a block of k_2 consecutive PBSs:*
1. $(1-\eta)$ fraction of queues are of type 0.
2. $2(1-\eta)$ fraction of queues are of type i for each $1 \leq i \leq \gamma$.
3. Remaining are either of type r (if $r-1 < \bar{J}$) or are connected to the MBS.

Here we present a brief sketch of the proof while the details are at the end of the chapter. We first show that the minimum distance between any two consecutive ON PBSs is $2r-1$. Let S_i be the set containing i-th ON PBS and its $r-1$ neighbours on each side (which are OFF). Each such set will contain a queue being served by the PBS of its own cell and two queues being served by PBSs i cells away $\forall\, 1 \leq i \leq \gamma$. Rest of the PBSs (excluding $\cup_i S_i$), if any, are either connected to the MBS or are of type r. Using this, we get equation (3.2) as the expression for minimum waiting time when η fraction of the PBSs have to be switched OFF.

3.3 Minimum Average Waiting Time

Theorem 5. *The minimum average waiting time for a fixed switch OFF ratio η is given by*

$$\overline{W}^*(\eta) = w(x) - (1-\eta)\left(\sum_{k=0}^{\min(r-1,\bar{J})} w(k) b_{r,k} + w(x)\left(1 + 2\min(r-1,\bar{J})\right)\right) \quad \textit{with} \tag{3.2}$$

$$x = \begin{cases} r & \textit{if } r-1 < \bar{J} \\ M & \textit{otherwise} \end{cases} \qquad b_{r,k} = \begin{cases} -1 & \textit{if } k = 0 \\ -2 & \textit{if } 1 \leq k \leq \min(r-1,\bar{J}) \end{cases}$$

when $h(r-1) < \eta \leq h(r)$ *with* $h(r) = \dfrac{2r}{1+2r}$.

We have the following results regarding the continuity and monotonicity properties of $\overline{W}^*(\eta)$.

Theorem 6. *The minimum waiting time (for a fixed switch OFF ratio) $\overline{W}^*(\eta)$ is a continuous function of η.*

Theorem 7. *The minimum waiting time (for a fixed switch OFF ratio) $\overline{W}^*(\eta)$ is an increasing function of η.*

3.4 Maximum Switch OFF Ratio to meet QoS Requirements

We have derived an expression for $\overline{W}^*(\eta)$ i.e. the minimum average waiting time possible for a given η. As seen from Theorem 6 and Theorem 7, $\overline{W}^*(\eta)$ is a continuous and increasing function of η. The average waiting time will be the least, i.e. $w(0)$, when all the PBSs are ON. It will be the maximum, i.e. $w(M)$, when all the PBSs are OFF and all the queues are being served by the MBS. Thus, average waiting time always takes values between $w(0)$ and $w(M)$. Hence, if $\overline{W}_{\text{QoS}} \geq w(M)$, then all PBSs can be switched OFF. If $\overline{W}_{\text{QoS}} < w(0)$, then it is not possible to meet the QoS requirement with the given system parameters.

When $w(0) \leq \overline{W}_{\text{QoS}} \leq w(M)$, the fraction η' which satisfies $\overline{W}^*(\eta') = \overline{W}_{\text{QoS}}$, is given by:

$$\eta' = 1 - \frac{w(x) - \overline{W}_{\text{QoS}}}{w(x)\left(1 + 2\min(r'-1, \bar{J}\right) + \displaystyle\sum_{k=0}^{\min(r'-1,\bar{J})} w(k) b_{r',k}}, \tag{3.3}$$

where r' is such that $\overline{W}^*(h(r'-1)) < \overline{W}_{\text{QoS}} \leq \overline{W}^*(h(r'))$.

Theorem 8. *The fraction η' is the solution to the optimization problem (2.3).*

Thus, η' is the maximum fraction of PBSs that can be switched OFF for the given load conditions (reflected through η's dependence on $w(i)$) while meeting the QoS constraint.

3.5 Future Work

We have assumed that resources are allocated for users of each PBS. It is possible that better service is achieved if a PBS which is serving users of multiple cells uses all its resources together for serving all the users. This analysis can also be done through simulation studies.

If interference is taken into account, the new waiting times will not fit directly into the same framework as before. We can look for alternate ways to extend this model to accommodate interference in the signals.

Further, other QoS parameters like blocking probability can be studied.

3.6 Proofs of Theorems and Lemmas

Proof of Lemma 1. Let us check if $s = \left\lceil p\frac{k_2}{k_1} \right\rceil - 1$ satisfies the equality. We have,

$$\left(p\frac{k_2}{k_1} - 1\right)\frac{k_1}{k_2} \leq \left(\left\lceil p\frac{k_2}{k_1}\right\rceil - 1\right)\frac{k_1}{k_2} < \left(p\frac{k_2}{k_1}\right)\frac{k_1}{k_2}.$$

$$\therefore p-1 \leq \left(\left\lceil p\frac{k_2}{k_1}\right\rceil - 1\right)\frac{k_1}{k_2} < p.$$

Hence, $\left\lfloor \left(\left\lceil p\frac{k_2}{k_1}\right\rceil - 1\right)\frac{k_1}{k_2}\right\rfloor = p-1$. As $\left\lfloor s\frac{k_1}{k_2}\right\rfloor$ is non-decreasing in s, it is less than p for all $s \leq \left\lceil p\frac{k_2}{k_1}\right\rceil - 1$.

Now, consider $s = \left\lceil p\frac{k_2}{k_1}\right\rceil$. We have,

$$p\frac{k_2}{k_1}\frac{k_1}{k_2} \leq \left\lceil p\frac{k_2}{k_1}\right\rceil\frac{k_1}{k_2} < \left(1 + p\frac{k_2}{k_1}\right)\frac{k_1}{k_2}.$$

$$\therefore p \leq \left\lceil p\frac{k_2}{k_1}\right\rceil\frac{k_1}{k_2} < p + \frac{k_1}{k_2}.$$

Hence, $\left\lfloor \left\lceil p\frac{k_2}{k_1}\right\rceil\frac{k_1}{k_2}\right\rfloor = p$. $\square$

Proof of Lemma 3. We want to find an expression for $l(\eta)$ in terms of η. Towards this, we proceed by proving the following:

1. $\{\eta : l(\eta) = r+1\} \subset \{\eta \leq h(r)\}$

2. $\{\eta \leq h(r)\} \subset \{\eta : l(\eta) \leq r+1\}$

3. $\{h(r-1) < \eta \leq h(r)\} = \{\eta : l(\eta) = r+1\}$.

Step 1: $\{\eta : l(\eta) = r+1\} \subset \{\eta \leq h(r)\}$

We first determine the permissible η for $l(\eta) = i \;\; \forall\, i$. Define

$$d_n(1) = \inf_{j>0}\left\{a^*_{n+j} = 1\right\} \text{ and}$$
$$d_n(k) = \inf_{j>d_n(k-1)}\left\{a^*_{n+j} = 1\right\}.$$

Note that, $d_n(k)$ represents the number of cells between the PBS in the nth cell and the kth next OFF PBS from it. Clearly, $d_n(k) \geq k$.

Let us consider the various possible values of $l(\eta)$.

1. $l(\eta) = 1$ means all users are being served at the rate θ^0. This is possible only when all the PBSs are ON. This can happen only when $\eta = 0$.

2. $l(\eta) = r + 1$ - This means that the distance of each user from its nearest ON PBS is among $0, 1, \cdots, r$ cells This will happen when the number of consecutive PBSs which are OFF is at most $2r$. This means for any OFF PBS at position $n, d_n(2r) > 2r$. In particular, this is true for $n = n_0$ (recall that $n_0 = mk_2$ and $a_{n_0} = 1$ i.e. the PBS located in the n_0 cell is OFF). Thus, from Lemma 2 $\forall\ p > 2r$

$$\left(n_0 + \left\lceil p\frac{k_2}{k_1}\right\rceil\right) - \left(n_0 + \left\lceil (p-2r)\frac{k_2}{k_1}\right\rceil\right) > 2r.$$
$$\therefore \left\lceil \frac{p}{\eta}\right\rceil - \left\lceil \frac{p-2r}{\eta}\right\rceil > 2r. \tag{3.4}$$

If possible let $\eta > h(r)$. Then,

$$-(2r+1) < -\frac{2}{\eta}r \leq -2r.$$
$$\therefore \left\lceil -\frac{2}{\eta}r\right\rceil = -2r.$$

Let us take $p = nk_1$ for some integer n. Then,

$$\left\lceil \frac{p}{\eta}\right\rceil - \left\lceil \frac{p-2r}{\eta}\right\rceil = nk_2 - (nk_2 + \left\lceil -\frac{2r}{\eta}\right\rceil) = 2r.$$

Thus, we have found a value of integer p for which equation (3.4) is not satisfied. Therefore, $l(\eta)$ cannot be $r+1$ for $\eta > h(r)$. Hence,

$$l(\eta) = r + 1 \implies \eta \leq h(r). \tag{3.5}$$

Step 2: $\{\eta \leq h(r)\} \subset \{\eta : l(\eta) \leq r + 1\}$

Now, let us consider the case when $\eta \leq \dfrac{2r}{1+2r}$ and check if it is possible to have $l(\eta) > r + 1$.

Assume, $l(\eta) \geq r + 2$. This implies that there exists two ON PBSs such that there are atleast $2r + 1$ consecutive OFF PBSs between them. Thus, for some q,

$$a_q = 0, a_{q+1} = 1, a_{q+2} = 1, \cdots a_{q+2r+1} = 1.$$

From Lemma 2, location of each OFF PBS can be written in the form $n_0 + \left\lceil \frac{z}{\eta}\right\rceil$ for some integer z. Thus, there exists a z such that,

$$q + 1 = n_0 + \left\lceil \frac{z}{\eta}\right\rceil,\ q + 2 = n_0 + \left\lceil \frac{z+1}{\eta}\right\rceil, \cdots,$$

$$q + 2r + 1 = n_0 + \left\lceil \frac{z + 2r}{\eta} \right\rceil.$$

Using the first and last equation, we get

$$\left\lceil \frac{z}{\eta} \right\rceil + 2r = \left\lceil \frac{z + 2r}{\eta} \right\rceil.$$

As ceiling of a number is strictly less than one more than the number,

$$\begin{aligned}
&\left\lceil \frac{z + 2r}{\eta} \right\rceil < \frac{z}{\eta} + 2r + 1, \\
&\Rightarrow \frac{z + 2r}{\eta} < \frac{z}{\eta} + 2r + 1, \\
&\Rightarrow \frac{2r}{\eta} < 2r + 1, \\
&\Rightarrow \eta > \frac{2r}{1 + 2r}.
\end{aligned}$$

This is a contradiction. Thus, $l(\eta) \leq r + 1$ for $\eta < h(r)$.

Step 3: $\{h(r-1) < \eta \leq h(r)\} = \{\eta : l(\eta) = r + 1\}$

Now consider $h(r-1) < \eta \leq h(r)$. We know that $l(\eta) \leq r + 1$. We want to find out the exact value of $l(\eta)$. Assume $l(\eta) \leq r = (r-1) + 1$. But from equation (3.5), $l(\eta) \leq (r-1) + 1$ implies $\eta \leq h(r-1)$. This is a contradiction. Thus, for $h(r-1) < \eta \leq h(r)$, the only possible value of $l(\eta)$ is $r + 1$. □

Proof of Theorem 4. Consider two consecutive ON PBSs. Let them be in cells q and $q+p$, where p is the number of cells between the two PBSs. Thus, we have, $a_q = a_{q+p} = 0$.
Let $h(r-1) < \eta \leq h(r)$ for some r. Thus, $l(\eta) = r + 1$.
Let $\lfloor q\eta \rfloor = \lfloor (q-1)\eta \rfloor = s$ and $\lfloor (q+p)\eta \rfloor = \lfloor (q+p-1)\eta \rfloor = s+o$. We have the following result -

Lemma 4. *If* $\lfloor q\eta \rfloor = \lfloor (q-1)\eta \rfloor = s$ *and* $\lfloor (q+p)\eta \rfloor = \lfloor (q+p-1)\eta \rfloor = s + o$*, then* $o \leq p - 1$.

Proof.

$$
\begin{aligned}
(q+p-1)\eta &= q\eta + (p-1)\eta, \\
&< q\eta + (p-1). \\
\therefore \lfloor (q+p-1)\eta \rfloor &\leq \lfloor q\eta + p - 1 \rfloor, \\
&= \lfloor q\eta \rfloor + p - 1. \\
\therefore s + o &\leq s + (p-1), \\
\Rightarrow o &\leq p - 1.
\end{aligned}
$$

We have $s \leq q\eta - \eta$ and $(q+p)\eta < s + o + 1$. This implies □

$$
\begin{aligned}
(q+p)\eta &< q\eta - \eta + o + 1, \\
\Rightarrow \quad \eta &< \frac{o+1}{1+p}, \\
\Rightarrow \quad \eta &< \frac{p}{1+p} \text{ (from Lemma 4).}
\end{aligned}
$$

It is easy to see that $\frac{i}{i+1}$ is an increasing function of i. We also know that $h(r-1) = \frac{2(r-1)}{1+2(r-1)} < \eta$. Thus, $p > 2(r-1)$ i.e. $p \geq 2r-1$. Thus, the minimum distance between two consecutive ON PBSs is $2r-1$ cells. This means that there are atleast $2r-2$ consecutive OFF PBSs between any two ON PBSs. Thus, if we define S_i to be the set containing the ith ON PBS and its $r-1$ neighbours on each side, then the sets S_is will be disjoint. This means that each set contains only one ON PBS.

Further, each set will contain one queue being served at the rate θ^0 (the queue in the cell of ON PBS) and two queues being served by PBS located i cells away, $1 \leq i \leq \gamma$. Thus, number of type i queues for $1 \leq i \leq \gamma$ is twice the number of type 0 queues i.e. twice the number of ON PBSs.

We know that in the activation vector $\mathbf{a}^*(\eta)$, the fraction of queues of type 0 i.e. fraction of ON PBSs is equal to $(1-\eta)$. As each ON PBS is associated to two queues of type i for $1 \leq i \leq \gamma$, the fraction of these type i queues is $2(1-\eta)$. If $\gamma = \bar{J}$, then the rest of the queues will be connected to the MBS. Whereas if $\gamma < \bar{J}$,then more queues can be connected to the PBSs. As $l(\eta) = r+1$, the only other type of queue possible is type r. Thus, the remaining queues are of type r.

□

Proof of Theorem 5. In the activation vector $\mathbf{a}^*(\eta)$, the fraction of OFF PBSs is exactly equal to η and hence in a block of k_2 PBSs, the number of OFF PBSs will be $\eta \times k_2 = k_1$.

Hence, the number of ON PBSs $= k_2 - k_1$. Thus if $r-1 < \bar{J}$, number of θ^i queues will be $2(k_2-k_1)$ for $1 \leq i \leq r-1$. Hence, number of θ^r queues is $k_2 - (1+2(r-1))(k_2-k_1) = k_1(2r-1) - 2k_2(r-1)$.

Therefore,

$$\begin{aligned}
\overline{W}^*(\eta) =& \frac{1}{k_2}\left((k_2-k_1)\, w(0) + 2\,(k_2-k_1)\, w(1) + \right.\\
& 2\,(k_2-k_1)\, w(2) + \cdots + 2\,(k_2-k_1)\, w(r-1) + \\
& \left. w(r)\,(k_1(2r-1) - 2k_2(r-1))\right), \\
=& (1-\eta)\, w(0) + 2\,(1-\eta)\, w(1) + \cdots + \\
& 2\,(1-\eta)\, w(r-1) + w(r)\,(\eta(2r-1) - 2(r-1)).
\end{aligned}$$

Similarly, when $r-1 \geq \bar{J}$, then number of type i queues will be $2(k_2-k_1)$ for $1 \leq i \leq \bar{J}$. Rest of the queues i.e. $k_1(2\bar{J}+1) - 2k_2\bar{J}$ queues will be connected to the MBS. Combining these two cases, we get equation (3.2) as the expression for minimum waiting time when η fraction of the PBSs have to be switched OFF. □

Proof of Theorem 6. We know that if $h(r-1) < \eta \leq h(r)$, then

1)If $r-1 < \bar{J}$,

$$\begin{aligned}
\overline{W}^*(\eta) = \quad & (1-\eta)\, w(0) + 2\,(1-\eta)\, w(1) + \cdots + \\
& 2\,(1-\eta)\, w(J) + w(r)\,(\eta(2r-1) - 2(r-1)). \qquad (3.6)
\end{aligned}$$

2) If $r-1 \geq \bar{J}$,

$$\begin{aligned}
\overline{W}^*(\eta) = \quad & (1-\eta)\, w(0) + 2\,(1-\eta)\, w(1) + \cdots + \\
& 2\,(1-\eta)\, w(r-1) + w(M)\left(\eta(2\bar{J}+1) - 2\bar{J}\right). \qquad (3.7)
\end{aligned}$$

We need to check if the waiting time is continuous at the boundaries i.e. for $\eta = h(r)$. For this value of η, we will find the right and left limits $(\overline{W}_r, \overline{W}_l)$. $\overline{W}_l$ will be evaluated based on the expression of $\overline{W}$ in the region $h(r-1) < \eta \leq h(r)$ whereas $\overline{W}_r$ will be evaluated based on the expression of $\overline{W}$ in the region $h(r) < \eta \leq h(r+1)$.

We can have 3 cases depending on relation between r and $\bar{J}$:

1. $r < \bar{J}$

   In this case, we will use equation (3.6) for evaluating both $\overline{W}_l$ and $\overline{W}_r$ (because

$r-1<\bar{J}$ and $r<\bar{J}$).

$$\begin{aligned}\overline{W}_l &= (1-\eta)\,w(0)+2\,(1-\eta)\,w(1)+\cdots+\\ &\quad 2\,(1-\eta)\,w(r-1)+w(r)\,(\eta(2r-1)-2(r-1))\,,\\ &= X+w(r)\,(\eta(2r-1)-2(r-1))\,,\\ &= X+\frac{2}{1+2r}w(r),\end{aligned}$$

where $X=(1-\eta)\,w(0)+2\,(1-\eta)\,w(1)+\cdots+2\,(1-\eta)\,w(r-1)$.

$$\begin{aligned}\overline{W}_r &= (1-\eta)\,w(0)+2\,(1-\eta)\,w(1)+\cdots+2\,(1-\eta)\,w(r-1)\\ &\quad +w(r)\,(\eta(2r-1)-2(r-1))+w(\theta^{r+1})\,(\eta(2(r+1)-1)-2r)\,,\\ &= X+w(r)\,(\eta(2r-1)-2(r-1))+w(r+1)\,(\eta(2(r+1)-1)-2(r))\,,\\ &= X+\frac{2}{1+2r}w(r)+w(r+1)\times 0=\ \overline{W}_l.\end{aligned}$$

Thus, in this case $\overline{W}^*(\eta)$ is continuous at the boundaries.

2. $r=\bar{J}$

   In this case, we will use (3.6) for evaluating $\overline{W_l}$ and (3.7) for evaluating $\overline{W_r}$ (because $r-1<\bar{J}$ and $r=\bar{J}$). As derived previously,

$$\overline{W}_l = X+\frac{2}{1+2r}w(r).$$

$$\begin{aligned}\overline{W}_r &= (1-\eta)\,w(0)+2\,(1-\eta)\,w(1)+\cdots+\\ &\quad 2\,(1-\eta)\,w(J)+w(M)\left(\eta(2\bar{J}+1)-2\bar{J}\right),\\ &= X+2\,(1-\eta)\,w(r)+w(M)\left(\eta(2\bar{J}+1)-2\bar{J}\right),\\ &= X+\frac{2}{1+2r}w(r)+w(M)\times 0=\overline{W}_l.\end{aligned}$$

   Thus, in this case also $\overline{W}^*(\eta)$ is continuous at the boundaries.

3. $r>\bar{J}$

   In this case, we will use (3.7) for evaluating both $\overline{W_l}$ and $\overline{W_r}$ (because $r-1\geq\bar{J}$

and $r \geq \bar{J}$). From (3.7),

$$\overline{W}^*(\eta) = (1-\eta)\, w(0) + 2\,(1-\eta)\, w(1) + \cdots +$$
$$2\,(1-\eta)\, w(J) + w(M)\left(\eta(2\bar{J}+1) - 2\bar{J}\right).$$

Clearly, this expression does not depend on r. Thus value of $\overline{W}^*(\eta)$ will be same using the expressions for $h(r-1) < \eta \leq h(r)$ and $h(r) < \eta \leq h(r+1)$. Hence, it will be continuous at the boundary i.e. when $\eta = h(r)$ for any integer r.

Therefore, in all three cases, we obtain continuity of waiting time at the boundaries. The continuity of waiting times at points other than the boundaries can be easily seen from the expression of $\overline{W}^*(\eta)$.

Thus, the waiting time is a continuous fucntion of η for all η. □

Proof of Theorem 7. Consider $h(r-1) < \eta \leq h(r)$ for some r. We examine the derivative of η in this interval.

Let us consider the case when $r - 1 < \bar{J}$.

$$\begin{aligned}
\frac{d\overline{W}^*(\eta)}{d\eta} &= \sum_{k=0}^{r} w(k) b_{r,k} + w(r-1)(1 + 2(r-1)),\\
&= -w(0) - 2w(1) - 2w(2) \cdots 2w(r-1) + (2r-1)w(r),\\
&> -w(0) - 2w(r) - 2w(r) \cdots 2w(r) + (2r-1)w(r),\\
&= -(2(r-1)+1)w(r) + (2r-1)w(r),\\
&= 0.
\end{aligned}$$

Thus, the derivative is positive. Hence, $\overline{W}^*(\eta)$ is an increasing function of η in the interval $h(r-1) < \eta \leq h(r), \forall\ r$. As $\overline{W}^*(\eta)$ is continuous at the interval boundaries i.e. at $\eta = h(r)$, we conclude that $\overline{W}^*(\eta)$ is an increasing function of $\eta, \eta \in [0, 1]$.

With similar arguments, we can prove the above even when $r - 1 \geq \bar{J}$. □

Proof of Theorem 8. If possible, assume there exist $\tilde{\eta} > \eta'$ and policies $(\mathbf{a}, J)$ such that

$$\tilde{\eta} \geq \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n \text{ and } \overline{W}(\mathbf{a}, J) \leq \overline{W}_{QoS}.$$

$\mathbf{a}^*(\tilde{\eta})$ represents the bracket sequence with switch OFF ratio $\tilde{\eta}$. By optimality of policies $(\mathbf{a}^*(\tilde{\eta}),\ \bar{J})$, we have

$$\overline{W}^*(\tilde{\eta}) = \overline{W}(\mathbf{a}^*(\tilde{\eta}), \bar{J}) \leq \overline{W}(\mathbf{a}, J) \leq \overline{W}_{QoS}.$$

On the other hand, by monotonicity of $\overline{W}^*(.)$ we have, $\overline{W}^*(\tilde{\eta}) > \overline{W}^*(\eta') = \overline{W}_{QoS}$. This is a contradiction. Thus, η' is the optimal value in (2.3) and $\mathbf{a}' := \mathbf{a}^*(\eta')$ and $J = \bar{J}$ are the optimal pair of policies solving the optimization problem (2.3). $\square$

Chapter 4

Decentralized Switching Policies

In the earlier chapters of this thesis, we have considered optimal switch OFF policies for a particular scenario - a linear deployment of pico base stations in a macro cell. In this chapter, we consider a single (stand-alone) pico base station and focus on its switching policies. The PBS may have been installed in areas to meet the demands of some periodic bursts of load.

This work considers a more general solution as it does not assume any particular deployment scenario.

4.1 System Model

We have a PBS installed in a macro cell. The PBS serves users [1] at a better rate as compared to the MBS. But at certain times, switching OFF the PBS is appropriate as the network QoS requirements can be met by the MBS itself. Thus, the PBS can be switched ON-OFF depending on the current traffic conditions and ensuring that the network QoS requirements are met.

As it is not feasible to switch ON-OFF base stations very quickly, the switching decision is taken at intervals of duration T (where T is large). Thus, the entire timeline is divided into slots of duration T and the switching decision is taken only once in every slot.

A very commonly used QoS parameter in cellular networks is average waiting time of a user. By Little's Law, the average waiting time can be equated (with a constant) to the

[1] By 'users' we refer to users in the vicinity of PBS i.e. all users in the pico cell.

average number of users waiting in the system (for a stable system). The average number of waiting users in the system over all the slots can be found by taking the integral of the number of waiting users present in the system and then dividing by the total duration.

Thus, the QoS metric in each slot is taken as the integral of the expected number of users in that slot. As we assume T to be large, for a stable system, this integral is equal to the expected number of waiting users times the length of the slot.

Here, we assume that when the PBS is ON, all users are being served by the PBS. Also, the user arrival process is a Poisson process with rate λ and the user file sizes are exponentially distributed. Like in the previous part, the BSs are modelled as M/G/1 queues. The service rate distribution will depend on the distribution of users within the pico cell, fading and shadowing parameters and the rates at which the MBS and PBS can offer service. (We are making no assumptions about any of these.)

4.1.1 Optimization Problem

Energy is spent i.e. cost is incurred in keeping the PBS ON. In addition to this, if the PBS is switched OFF, users might experience a poor QoS which we model as an inconvenience cost . The total cost in any slot is given by the sum of these two. Let C be the cost of keeping the PBS ON per unit time and $W_k(a_k)$ represent the value of the QoS metric (user inconvenience cost) in the k-th slot with a_k representing the ON/OFF status of the PBS in the k-th slot. We define the activation vector $\mathbf{a} = (a_1, a_2, \cdots, a_N)$ with $a_i = 1(0)$ signifying that the PBS is ON (OFF) in the i-th slot.

Our aim is to find the activation vector that minimizes the average cost i.e.

$$\min_{\mathbf{a}} \left(CT \frac{1}{N} \sum_{k=1}^{N} a_k + \frac{1}{N} \sum_{k=1}^{N} W_k(a_k) \right).$$

We denote the average cost by $J(\mathbf{a})$ when the activation vector $\mathbf{a}$ is used.

The system is stable when the arrival rate of users is less than the average service rate. If the average service rates of the MBS and PBS are μ_m and μ_p respectively, then we expect different system behaviours when $\lambda < \mu_m < \mu_p$, $\mu_m < \lambda < \mu_p$ and $\mu_m < \mu_p < \lambda$. In the third case, the system will not be stable for any ON-OFF policy, Thus, the number of users in the system will keep on building up irrespective of the activation policy. Hence, only the first two cases are of interest.

4.2 Case when $\lambda < \mu_m < \mu_p$

As the arrival rate is less than the service rates of both the PBS and the MBS, the queue would be stable in both cases - when the users are being served by the PBS and when the users are being served by the MBS. Since the time slots are large and the queue is stable, the inconvenience cost will be approximately equal to the stationary value of the average number of waiting users times the duration of the slot. This value can be easily obtained from M/G/1 queue related results.

For an M/G/1 queue, the expected number of waiting users in the system is given by $\dfrac{\lambda^2 E[B^2]}{2\,(1-\lambda E[B])}$ where B represents the service time. Let $B^m(B^p)$ denote the random service time when a user is being served by the MBS (PBS) with mean $\frac{1}{\mu_m}\left(\frac{1}{\mu_p}\right)$. Then, the user inconvenience cost when MBS (PBS) is serving the users is given by $\dfrac{T\lambda^2 E[(B^m)^2]}{2\,(1-\lambda E[B^m])}\left(\dfrac{T\lambda^2 E[(B^p)^2]}{2\,(1-\lambda E[B^p])}\right)$. Thus,

$$\begin{aligned} W_k(a_k) =& a_k \frac{T\lambda^2 E[(B^p)^2]}{2\,(1-\lambda E[B^p])} + (1-a_k)\frac{T\lambda^2 E[(B^m)^2]}{2\,(1-\lambda E[B^m])} \\ =& \frac{T\lambda^2 E[(B^m)^2]}{2\,(1-\lambda E[B^m])} + a_k\left(\frac{T\lambda^2 E[(B^p)^2]}{2\,(1-\lambda E[B^p])} - \frac{T\lambda^2 E[(B^m)^2]}{2\,(1-\lambda E[B^m])}\right). \end{aligned}$$

4.2.1 Fixed Arrival Rates in All Slots

We first consider the scenario with the arrival rate λ fixed for the entire duration.

Here the average cost is

$$\begin{aligned} J(\mathbf{a}) &= CT\frac{1}{N}\sum_{k=1}^{N} a_k + \frac{1}{N}\sum_{k=1}^{N} W_k(a_k), \\ &= \frac{1}{N}\sum_{k=1}^{N}\left(CTa_k + \frac{T\lambda^2 E[(B^m)^2]}{2\,(1-\lambda E[B^m])} + a_k\left(\frac{T\lambda^2 E[(B^p)^2]}{2\,(1-\lambda E[B^p])} - \frac{T\lambda^2 E[(B^m)^2]}{2\,(1-\lambda E[B^m])}\right)\right), \\ &= \frac{T\lambda^2 E[(B^m)^2]}{2\,(1-\lambda E[B^m])} + \frac{1}{N}\sum_{k=1}^{N} a_k\left(CT + \frac{T\lambda^2 E[(B^p)^2]}{2\,(1-\lambda E[B^p])} - \frac{T\lambda^2 E[(B^m)^2]}{2\,(1-\lambda E[B^m])}\right). \end{aligned}$$

Note that $J(\mathbf{a})$ varies monotonically with $\frac{1}{N}\sum_{k=1}^{N} a_k$. Thus, the optimal policy i.e. the policy with minimum cost is given by

$$a_k^* = \begin{cases} 0 \text{ if } \left(CT + \dfrac{T\lambda^2 E[(B^p)^2]}{2\,(1-\lambda E[B^p])} - \dfrac{T\lambda^2 E[(B^m)^2]}{2\,(1-\lambda E[B^m])}\right) > 0 \\ 1 \text{ otherwise} \end{cases}$$

$\forall\ k$.

4.2.2 Different Arrival Rates in Different Slots

It is less likely that the arrival rate of users will remain fixed during all the slots. So we now consider the arrival rate of users to be variable i.e. λ_i in slot i. For cellular networks, the traffic profile has a periodic behaviour. Thus, we assume that the arrival rates in different slots can be estimated from previous data.

As the slot duration is large, the number of users initially present in the system will not affect the average waiting time. Further, the ON/OFF status of the PBS in one slot does not affect the waiting time of the users in the next slot.

MDP Formulation

We formulate this problem as a finite horizon Markov Decision Process (MDP) problem. The horizon is finite as any day will have only finite number of slots and the traffic patterns have generally been observed to be similar for weekdays (and another pattern for weekends).

MDP corresponding to our system model is characterized by the following-

- The arrival rate represents the state (X) of the system i.e. state space $S = (0, \mu_P)$

- There are two possible actions (Y) - switch ON the PBS and switch OFF the PBS represented by 1 and 0 respectively. Thus, $A = \{0, 1\}$.

- The decision epochs are the beginning of every new slot i.e. time instants $\{0, T, 2T, \cdots\}$

- The transition probabilities are given by the stochastic process governing the arrival rates (due to the fact that action taken in a slot does not affect user waiting time in the next slot).

- The cost in each slot is given by $R_{a_k}(x) = CTa_k + W_k(a_k)$.

As discussed before, the switching ON/OFF action in one slot does not affect the average number of waiting users in the next slot and hence the cost in the next slot.

Moreover, the state (arrival rate) in the next slot is unaffected by the action in the current slot. Thus, the optimal policy would be a myopic policy i.e. minimize the cost in each slot separately which is given by -

$$a_k^* = \begin{cases} 1 \text{ if } CT + \dfrac{T{\lambda_k}^2 E[(B^p)^2]}{2\,(1-\lambda_k E[B^p])} < \dfrac{T{\lambda_k}^2 E[(B^m)^2]}{2\,(1-\lambda_k E[B^m])} \\ 0 \text{ otherwise.} \end{cases}$$

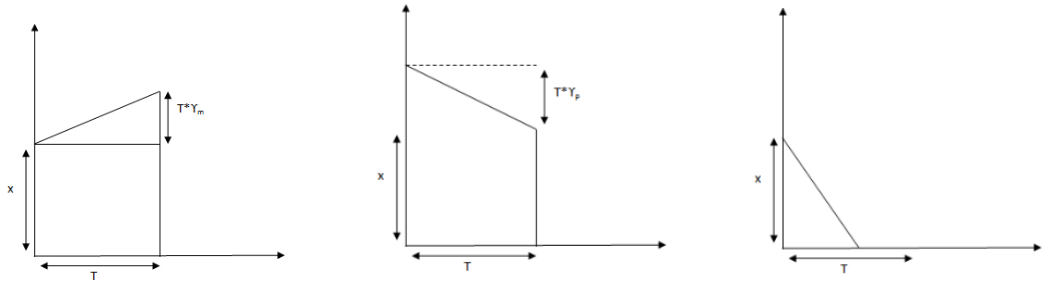
4.3 Case when $\mu_m < \lambda < \mu_p$

The analysis here differs from that of the previous case because the system would not be stable when the PBS is switched OFF and the MBS is serving users. Thus, the average number of waiting users in the system can not be expressed by $\dfrac{T\lambda^2 E[(B^p)^2]}{2\,(1-\lambda E[B^p])}$. For analysing this situation, we use a fluid model for the queuing system. In this model, the arrivals and departures happen at the mean rate throughout. This model is a fair approximation when λ and μ i.e. the arrival rate and service rate are very large. In our system, if we consider the jobs in terms of packets, then both the arrival and departure rates will be very high (the waiting time will then correspond to the end-to-end packet delay). Thus, this model would be appropriate.

We study this case with constant λ in all slots.

The MDP formulation is slightly different from the previous case. Here, we define the number of users at the beginning of the slot as the state of the system. Note that, in the fluid model, the state can take continuous values (and not just discrete numbers). Thus, the state space of the MDP is continuous.

In this model, the user arrival takes place at the constant rate of λ and the service happens at the constant rate of $\mu_m(\mu_p)$ when the MBS (PBS) is serving the users. Thus, the number of users in the system increases (decreases) at the rate $\gamma_m = \lambda - \mu_m(\gamma_p = \lambda - \mu_p)$ when the MBS (PBS) is serving the user. We find the integral of the number of users by calculating the area under this straight line. When the PBS is serving the users, the number of users in the system keeps on decreasing. If the number of users at the beginning of the slot is less than $\gamma_P T$, the number of users will fall to 0 during the slot. After that the number of users remains at 0 as the arrival rate of users is less than the service rate. Thus, the immediate cost $R_a(x)$ i.e. the cost when action a is taken in state x is given by -



(a) $a = 0$ (b) $a = 1, x \geq \gamma_p T$ (c) $a = 1, x < \gamma_p T$

Figure 4.1: User Inconvenience Cost

$$R_a(x) = \begin{cases} xT + 1/2\gamma_m T^2 & \text{if } a = 0 \\ xT - 1/2\gamma_p T^2 + CT & \text{if } a = 1 \text{ and } \gamma_p T \leq x \\ x^2/(2\gamma_p) + CT & \text{if } a = 1 \text{ and } \gamma_p T > x, \end{cases}$$

where $\gamma_p = \mu_P - \lambda$ and $\gamma_m = \lambda - \mu_M$. This can be understood from the area under the curve of number of users in the system (Figure 4.1).

This cost function satisfies the following property -

Lemma 5. *For $X_1 > X_2$,*

$$R_Y(X_1) - R_Y(X_2) \begin{cases} = T(X_1 - X_2) & \textit{if } \; Y = \gamma_m \\ \leq T(X_1 - X_2) & \textit{otherwise} \end{cases}$$

If the PBS is switched ON, then the number of users in the system decreases at the rate $(\mu_p - \lambda)$ whereas if the PBS is switched OFF i.e. MBS is serving the users then the number of users increases at the rate $(\lambda - \mu_m)$. Thus, the transition probabilities are given as follows -

$$P(X_{n+1} = y | X_n = x, a_n = 0) = \begin{cases} 1 & \text{if } y = x + \gamma_m T \\ 0 & \text{otherwise} \end{cases} \quad \text{and}$$

$$P(X_{n+1} = y | X_n = x, a_n = 1) = \begin{cases} 1 & \text{if } y = (x - \gamma_p T)^+ \\ 0 & \text{otherwise.} \end{cases}$$

The decision epochs and action space remain same as before.

To find the optimal activation vector, we use dynamic programming. Let $V^n(x)$ represent the optimal total cost in slots n to N when the state in slot n is x. Thus,

$V^{N+1}(x) := 0 \ \forall x$ and

$$V^n(x) = \min_{a \in \{0,1\}} \left\{ R_a(x) + V^{n+1}(f(x,a)) \right\} \tag{4.1}$$

where $f(x,0) = x + \gamma_m T$ and $f(x,1) = (x - \gamma_p T)^+$.

4.3.1 Properties of $V^n(x)$

Before we derive the optimal policies, it is useful to understand the properties of the value function $V^n(x)$. We obtain the following results -

Theorem 9. *$V^n(x)$ is monotonically increasing in x for a fixed n.*

Theorem 10. *$V^n(x)$ is continuous in x for a fixed n and for $x' > x$*

$$V^n(x') - V^n(x) \leq T(N - n + 1)(x' - x).$$

The proofs of these two theorems follow the proof outline in ([18]) and are given in section 4.5.

4.3.2 Optimal policy

Let $a_k^*(x)$ represent the optimal policy in the kth slot.

The form of the optimal policy depends on the relation between the parameters C, T, γ_m and γ_p. We begin by considering the case $C < \frac{(\gamma_p + \gamma_m)T}{2}$ and solve for $V^N(x)$, $V^{N-1}(x)$ and then study the observed patterns.

As $V^{N+1}(x) = 0$,

$$V^N(x) = \min_{a \in \{0,1\}} \{R_a(x)\}. \tag{4.2}$$

As $R_1(x)$ is defined piecewise, we have two cases -

- $\gamma_p T \leq x$

$$\begin{aligned} V^N(x) &= \min \left\{ (xT + 1/2\gamma_m T^2), (xT - 1/2\gamma_p T^2 + CT) \right\}, \\ &= xT - 1/2\gamma_p T^2 + CT. \end{aligned}$$

Thus, the optimal decision at this epoch is $a_N^*(x) = 1 \ \forall \ \gamma_p T \leq x$.

- $\gamma_p T > x$

$$V^N(x) = \min\left\{(xT + 1/2\gamma_m T^2), x^2/(2\gamma_p) + CT)\right\}. \tag{4.3}$$

Thus, $V^N(x) = \begin{cases} xT + 1/2\gamma_m T^2 & \text{if } x \le x_N^* \\ x^2/(2\gamma_p) + CT & \text{if } x > x_N^*, \end{cases}$

where $x_N^* = T\gamma_p - \sqrt{(T^2\gamma_p^2 - 2T\gamma_p C + \gamma_p\gamma_m T^2)}$.

Thus, in this case $a_N^*(x) = 0$ when $x < x_N^*$ and 1 otherwise.

Combining the above, we get

$$V^N(x) = \begin{cases} xT + 1/2\gamma_m T^2 & \text{if } x \le x_N^* \\ x^2/(2\gamma_p) + CT & \text{if } x_N^* \le x \le \gamma_p T \\ xT + 1/2\gamma_m T^2 & \text{if } \gamma_p T \le x. \end{cases} \quad \text{and } a_N^*(x) = \begin{cases} 0 \text{ if } x < x_N^* \\ 1 \text{ if } x \ge x_N^*. \end{cases}$$

Thus, the optimal policy for the Nth slot is a threshold policy with threshold x_N^*.

Using backward induction, we can similarly obtain the optimal policy for the $(N-1)$-th slot as

$$a_{N-1}^*(x) = \begin{cases} 0 & \text{if } x < x_{N-1}^* \\ 1 & \text{if } x \ge x_{N-1}^* \end{cases},$$

where $x_{N-1}^* = 2\gamma_p T - \sqrt{4\gamma_p T^2 - 2CT\gamma_p + 3\gamma_m\gamma_p T^2}$ and

$$V^{N-1}(x) = \begin{cases} 2Tx + 2\gamma_m T^2 & \text{if } x \le x_{N-1}^* \\ x^2/2\gamma_p + CT^2 + \gamma_m T^2 & \text{if } x_{N-1}^* < x \le \gamma_p T \\ 2Tx - \frac{3}{2}\gamma_p T^2 + \gamma_m \frac{T^2}{2} + CT & \text{if } \gamma_p T < x \le \gamma_p T + x_N^* \\ \frac{T}{2}(2x - \gamma_p T) + CT + \frac{(x-\gamma_p T)^2}{2\gamma_p} + CT & \text{if } \gamma_p T + x_N^* < x \le 2\gamma_p T \\ 2Tx - 2T^2\gamma_p + 2CT & \text{if } x > 2\gamma_p T. \end{cases}$$

Thus, for the $(N-1)$-th slot also, the optimal policy turns to be a threshold policy. Therefore, it is likely that the optimal policy is a threshold policy for other slots as well when $C < \frac{T}{2}(\gamma_m + \gamma_p)$.

Let $\Gamma_y(x)$ denote a policy with threshold y i.e.

$$\Gamma_y(x) = \begin{cases} 0 \text{ if } x < y \\ 1 \text{ if } x \ge y. \end{cases}$$

For the general case $C > (2k+1)\frac{T}{2}(\gamma_m + \gamma_p)$ with $k \geq 0$ we obtain the following result (the case of $C < \frac{T}{2}(\gamma_m + \gamma_p)$ will be covered in Theorem 12) :

Theorem 11. *If $C > (2k+1)\frac{T}{2}(\gamma_m + \gamma_p)$, then $a^*_{N-j}(x) = 0 \; \forall \; x \; , \; 0 \leq j \leq k$ where $k \geq 0$.*

Corollary: If $C > (2N-1)\frac{T}{2}(\gamma_m + \gamma_p)$, then $a^*_k(x) = 0 \; \forall \; x, k$.

The proof of this theorem is given in section 4.5.

This result is intuitive as we would expect that if the cost of keeping the PBS ON is very high, then it would be better to keep the PBS switched OFF throughout. The above result gives the precise value of this cost threshold.

The previous theorem has described the optimal action for slots $N, N-1, \cdots N-k$. We need to derive the optimal actions in the slots before $N-k$. Towards this, we have the following result-

Lemma 6. *For $k \geq 0$, if $(2k-1)\frac{T}{2}(\gamma_m + \gamma_p) < C < (2k+1)\frac{T}{2}(\gamma_m + \gamma_p)$, then $a^*_{N-k}(x) = \Gamma_{x^*_{N-k}}(x)$ where $x^*_{N-k} = (k+1)\gamma_p T - \sqrt{\gamma_p T^2 \left((k+1)^2\gamma_p + (2k+1)\gamma_m - \frac{2C}{T}\right)} < \gamma_p T$.*

Finally, the optimal actions in all the slots before $N-k$ can be given as follows (proof in section 4.5)-

Theorem 12. *If $(2k-1)\frac{T}{2}(\gamma_m + \gamma_p) < C < (2k+1)\frac{T}{2}(\gamma_m + \gamma_p)$, then there exists x^*_s for $1 \leq s \leq N-k$ such that $0 \leq x^*_s < \gamma_p T$ and*

$$a^*_s(x) = \begin{cases} \Gamma_{x^*_s}(x) & \text{when } 1 \leq s \leq N-k \\ 0 & \text{when } N-k < s \leq N. \end{cases}$$

Thus, we have shown that the optimal policy in each slot is either to switch OFF the PBS for all x or is a threshold policy with the threshold value less than $\gamma_p T$.

4.4 Future Work

We have derived the optimal policy when the arrival rate is fixed throughout. It would be interesting to study the optimal policy when the arrival rates are varying. For example, we can consider a Markov Modulated Arrival process (MMAP). For cellular networks, the user traffic generally has cyclostationary behaviour. This means that the expected arrival rate in any slot can be approximated from the data of previous slots (in previous days) but some randomness can also be expected. Thus, the effect of some random variations in the arrival rate even within a slot can be incorporated in future study.

4.5 Proofs of Theorems and Lemmas

Proof of Lemma 5. Let $S = R_Y(X_1) - R_Y(X_2)$.

1. $Y = \gamma_m$
$$\begin{aligned} S &= \frac{T}{2}(2X_1 + \gamma_m T) - \frac{T}{2}(2X_2 + \gamma_m T) \\ &= T(X_1 - X_2). \end{aligned}$$

2. $Y = \gamma_p$

   (a) $X_1, X_2 \leq \gamma_p T$
$$\begin{aligned} S &= \left(\frac{X_1^2}{2\gamma_p} + CT\right) - \left(\frac{X_2^2}{2\gamma_p} + CT\right) \\ &= \frac{(X_1 - X_2)(X_1 + X_2)}{2\gamma_p}, \\ &\leq \frac{(X_1 - X_2)}{2\gamma_p} 2\gamma_p T \qquad (\because (X_1 + X_2) < 2\gamma_p T) \\ &= (X_1 - X_2)T. \end{aligned}$$

   (b) $\gamma_p T < X_2 < X_1$
$$\begin{aligned} S &= \left(TX_1 - \frac{\gamma_p T^2}{2} + CT\right) - \left(TX_2 - \frac{\gamma_p T^2}{2} + CT\right) \\ &= T(X_1 - X_2). \end{aligned}$$

   (c) $X_2 \leq \gamma_p T < X_1$

Let $X_1 = \gamma_p T + \delta_1$ and $X_2 = \gamma_p T - \delta_2$ where $\delta_1 > 0$ and $0 < \delta 2 < \gamma_p T$.

$$\begin{aligned}
S &= \left(T(\gamma_p T + \delta_1) - \frac{\gamma_p T^2}{2} + CT\right) - \left(\frac{(\gamma_p T - \delta_2)^2}{2\gamma_p} + CT\right) \\
&= \delta_1 T + \delta_2 T - \frac{\delta_2^2}{2\gamma_p} \\
&= \delta_1 T + \frac{\delta_2}{2\gamma_p}(2\gamma_p T - \delta_2), \\
&\leq \delta_1 T + \frac{\delta_2}{2\gamma_p} 2\gamma_p T \\
&= T(\delta_1 + \delta_2) \\
&= T(X_1 - X_2).
\end{aligned}$$

□

Proof of Theorem 9. Let $x < x'$. x and x' represent the state in the nth slot.

Let $< Y_k >$ be the optimal activation vector for states $< X'_k >$ and arrival rate λ. We have $X'_n = x'$ and define inductively,

$$X'_{k+1} = \left(X'_k - Y_k T + \lambda T\right)^+, k > n.$$

Similarly, we take $X_n = x$ and inductively define the process $< X_k >$ as

$$X_{k+1} = \left(X_k - Y_k T + \lambda T\right)^+, k > n.$$

It can be shown that $X_k \leq X'_k \ \forall k$.

Let $J^n_{<Y_k>}(x)$ be the cost incurred from slot n to N when the actions in each slot is given by the sequence $< Y_k >$. Now,

$$\begin{aligned}
V^n(x) &\leq J^n_{<Y_k>}(x) \\
&= E\left(\sum_{i=n}^{N} R_{Y_i}(X_i)\right) \\
&\leq E\left(\sum_{i=n}^{N} R_{Y_i}(X'_i)\right) \quad (R_y(x) \text{ monotonically increases in } x.) \\
&= V_n(x').
\end{aligned}$$

□

Proof of 10. Let $x < x'$. Let $< Y_k >$ be the optimal activation vector for states $< X_k >$ and arrival rate λ. We take $X_n = x$ and define inductively,

$$X_{k+1} = \left(X_k - Y_k T + \lambda T\right)^+, k > n.$$

Similarly, we take $X'_n = x'$ and inductively define the process $< X'_k >$ as

$$X'_{k+1} = \left(X'_k - Y_k T + \lambda T\right)^+, k > n.$$

It can be shown that $X_k \le X'_k \ \forall k$.

$$\begin{aligned}
V^n(x') - V^n(x) &\le J^n_{\langle Y_k \rangle}(x') - V^n(x) \\
&= E\left(\sum_{i=n}^{N} \left(R_{Y_i}(X'_i) - R_{Y_i}(X_i)\right)\right), \\
&\le E\left(\sum_{i=n}^{N} \left(T(X'_i - X_i)\right)\right) \text{(Lemma 5)}, \\
&\le T(N-n+1)(x'-x) \quad (\because X'_k - X_k \le x' - x).
\end{aligned}$$

This proves the continuity of $V^n(x)$ and also derives the required bound. □

Proof of 11. Consider the case $k = 0$ i.e. $C > \frac{T}{2}(\gamma_m + \gamma_p)$. It can be easily checked that $a^*_N(x) = 0 \ \forall x$. Assume that the above claim holds for $k = 1, 2, \cdots, j-1$. We will prove the claim for $k = j$.

We have, $C > (2j+1)\frac{T}{2}(\gamma_m + \gamma_p)$. As $C > (2l+1)\frac{T}{2}(\gamma_m + \gamma_p)$, $l < j$, from the assumption we have $a^*_{N-l}(x) = 0 \ \forall x, \ 0 \le l < j$. Thus, we only have to prove $a^*_{N-j}(x) = 0 \ \forall x$.

Consider $x > \gamma_p T$, the optimal action would be to switch OFF the PBS if

$$\begin{aligned}
& R_0(x) + V^{N-j+1}(x + \gamma_m T) &<& \quad R_1(x) + V^{N-j+1}(x - \gamma_p T) \\
\iff & \frac{T}{2}(j+1)(2x + (j+1)\gamma_m T) &<& \quad CT + Tx - \gamma_p \frac{T^2}{2} + \frac{jT}{2}(2x - 2\gamma_p T + j\gamma_m T) \\
\iff & \frac{2j+1}{2} T(\gamma_m + \gamma_p) &<& \quad C,
\end{aligned}$$

which is true. Thus, $a^*_{N-j}(x) = 0 \ \forall \ x > \gamma_p T$.

Now consider $x < \gamma_p T$. The optimal action would be to switch OFF the PBS if

$$\begin{aligned}
& R_0(x) + V^{N-j+1}(x + \gamma_m T) &<& \quad R_1(x) + V^{N-j+1}(x - \gamma_p T) \\
\iff & \frac{(j+1)T}{2}(2x + (j+1)\gamma_m T) &<& \quad CT + \frac{x^2}{2\gamma_p} + \frac{1}{2} j^2 T^2 \gamma_m \\
\iff & \frac{(j+1)^2 \gamma_p T^2}{2} - CT + \frac{(2j+1)\gamma_m T^2}{2} &<& \quad \frac{(\gamma_p T(j+1) - x)^2}{2\gamma_p}.
\end{aligned}$$

In the last inequality, the right hand side term(RHS) is a decreasing function of x for $x < \gamma_p T$ whereas the left hand side term(LHS) is a constant. Thus, if the RHS > LHS for $x = \gamma_p T$, then RHS > LHS for all x in the considered region ($x < \gamma_p T$).

Value of RHS at $x = \gamma_p T$ is $\dfrac{j^2T^2\gamma_p}{2}$ which is greater than the LHS $\left(\text{ as } \dfrac{2j+1}{2}T(\gamma_m+\gamma_p) < C\right)$.

Thus, even for $x < \gamma_p T$, $a^*_{N-j}(x) = 0$.

Combining the above two cases, we get $a^*_{N-j}(x) = 0\ \forall x$. Hence, by induction we have shown that if $C > (2k+1)\frac{T}{2}(\gamma_m+\gamma_p)$ then $a^*_{N-k}(x) = 0\ \forall\ x\ ,\ 0 \leqslant k \leqslant N$. □

Proof of Lemma 6. As $(2k-1)\frac{T}{2}(\gamma_m+\gamma_p) < C$, $a^*_N(x) = a^*_{N-1}(x) = \cdots a^*_{N-(k-1)}(x) = 0\ \forall x$. Thus, $V^{N-(k-1)}(x) = \frac{kT}{2}(2x + k\gamma_m T)$.

For $x > \gamma_p T$, the optimal policy would be to switch ON the PBS if,

$$\begin{aligned}
& R_1(x) + V^{N-k+1}(x-\gamma_p T) &<& \quad R_0(x) + V^{N-k+1}(x+\gamma_m T) \\
\iff\ & \tfrac{T}{2}(2x-\gamma_p T) + V^{N-(k-1)}(x-\gamma_p T) + CT &<& \quad \tfrac{T}{2}(2x+\gamma_m T) + V^{N-(k-1)}(x+\gamma_m T) \\
\iff\ & -\tfrac{\gamma_p T^2}{2} - \tfrac{\gamma_m}{T^2}/2 + CT - kT^2(\gamma_p+\gamma_m) &<& \quad 0 \\
\iff\ & C &<& \quad \tfrac{2k+1}{2}(\gamma_p+\gamma_m)T.
\end{aligned}$$

which is true. Thus, $a^*_{N-k}(x) = 1$ for $x > \gamma_p T$.

For $x < \gamma_p T$, the optimal policy would be to switch OFF the PBS if,

$$\begin{aligned}
& R_0(x) + V^{N-k+1}(x+\gamma_m T) &<& \quad R_1(x) + V^{N-k+1}(x-\gamma_p T) \\
\iff\ & \tfrac{T}{2}(2x+\gamma_m T) + V^{N-(k-1)}(x+\gamma_m T) &<& \quad \tfrac{x^2}{2\gamma_p} + CT + V^{N-(k-1)}(0) \\
\iff\ & \tfrac{(\gamma_p T - x)^2}{2\gamma_p} - \tfrac{\gamma_p T^2}{2} - \tfrac{\gamma_m T^2}{2} - kT(x+\gamma_m T) + CT &>& \quad 0 \\
\iff\ & x &<& \quad x^*_{N-k},
\end{aligned}$$

where, $x^*_{N-k} = \left((k+1)\gamma_p T - \sqrt{\gamma_p T^2\left((k+1)^2\gamma_p + (2k+1)\gamma_m - \frac{2C}{T}\right)}\right)^+$.

Combining with the case when $x > \gamma_p T$, we get

$$a^*_{N-k}(x) = \begin{cases} 0 \text{ for } 0 \le x < x^*_{N-k} \\ 1 \text{ if } x^*_{N-k} \le x. \end{cases}$$

Thus, the optimal policy is a threshold policy with the threshold x^*_{N-k}.

Further, we want $x^*_{N-k} < \gamma_p T$. This is trivially true if $(k+1)\gamma_p T - \sqrt{\gamma_p T^2\left((k+1)^2\gamma_p + (2k+1)\gamma_m - \frac{2C}{T}\right)} < 0$. When it is greater than 0, we want

$$\begin{aligned}
& & x^*_{N-k} &< \gamma_p T \\
&\iff & (k+1)\gamma_p T - \sqrt{\gamma_p T^2\left((k+1)^2\gamma_p + (2k+1)\gamma_m - \tfrac{2C}{T}\right)} &< \gamma_p T \\
&\iff & (k+1)^2\gamma_p^2 + (2k+1)\gamma_m\gamma_p - \tfrac{2C}{T\gamma_p} &> k^2\gamma_p^2 \\
&\iff & C &< \tfrac{(2k+1)}{2}(\gamma_p+\gamma_m)T,
\end{aligned}$$

which is true. Hence, $x^*_{N-k} < \gamma_p T$. □

Proof of Theorem 12. From Theorem 11, $a^*_s(x) = 0\ \forall x, N-k < s \leq N$. For $1 \leq s \leq N-k$, we will prove by induction. From Lemma 6, the optimal policy in slot $N-k$ is a threshold policy with threshold $x^*_{N-k} < \gamma_p T$.

Now, assume that $a^*_l(x) = Th_{x^*_l}(x)$ with $x^*_l < \gamma_p T$ for $j < l \leq N-k$. We will prove that the above also holds for $l = j$ (reverse induction).

1) Consider $x < \gamma_p T$. The optimal policy in slot j will be to switch OFF if,

$$\begin{aligned}
& & R_0(x) + V^{j+1}(x+\gamma_m T) &< R_1(x) + V^{j+1}(0) \\
&\iff & Tx + \tfrac{\gamma_m T^2}{2} + V^{j+1}(x+\gamma_m T) &< \tfrac{x^2}{2\gamma_p} + V^{j+1}(0) + CT \\
&\iff & \tfrac{(2\gamma_p T - x)^2}{2\gamma_p} + V^{j+1}(0) - V^{j+1}(x+\gamma_m T) + CT - \tfrac{\gamma_m T^2}{2} - \tfrac{\gamma_p T^2}{2} &> 0.
\end{aligned}$$

Let $h(x) = \frac{(\gamma_p T - x)^2}{2\gamma_p} + V^{j+1}(0) - V^{j+1}(x+\gamma_m T) + CT - \frac{\gamma_m T^2}{2} - \frac{\gamma_p T^2}{2}$. $h(x)$ is a decreasing function of x.

$$h(0) = CT - \frac{\gamma_m T^2}{2} + V^{j+1}(0) - V^{j+1}(\gamma_m T).$$

$$\begin{aligned}
h(\gamma_p T) =& V^{j+1}(0) - V^{j+1}(\gamma_p T + \gamma_m T) + CT - \frac{\gamma_m T^2}{2} - \frac{\gamma_p T^2}{2} \\
=& \min\left(\frac{\gamma_m T^2}{2} + V^{j+2}(\gamma_m T), CT - \frac{\gamma_p T^2}{2} + V^{j+2}(0)\right) - \left(\gamma_m T^2 + \frac{\gamma_p T^2}{2} + \right. \\
& \left. CT + V^{j+2}(\gamma_m T)\right) + CT - \frac{\gamma_m T^2}{2} - \frac{\gamma_p T^2}{2} \\
\leq& -T^2(\gamma_m + \gamma_p), \\
<& \ 0.
\end{aligned}$$

Thus, if $h(0) < 0$, then $a^*_j(x) = 1\ \forall x < \gamma_p T$. In this case, define $x^*_j = 0$.
If $h(0) \geq 0$, then there exists $x^*(j) < \gamma_p T$ such that $h(x^*_j) = 0$ (because $h(x)$ is a decreasing function for $x < \gamma_p T$ and $h(\gamma_p T) < 0$).

Thus, $a^*_j(x) = \begin{cases} 0 \text{ if } x < x^*_j \\ 1 \text{ if } x^*_j \leq x < \gamma_p T. \end{cases}$

2) Consider $l\gamma_p T < x < (l+1)\gamma_p T$ where $l \geq 1$. There are two possibilities -

- $j + l > N - k$

  The optimal policy would be to switch ON the PBS if -
  $R_1(x) + R_1(x - \gamma_p T) + \cdots + R_1(x - (z-1)\gamma_p T) + V^{N-k+1}(x - z\gamma_p T) <$
  $R_0(x) + R_1(x + \gamma_m T) + R_1(x + \gamma_m T - \gamma_p T) + \cdots + R_1(x + \gamma_m T - (z-2)\gamma_p T) +$
  $V^{N-k+1}(x + \gamma_m T - (z-1)\gamma_p T)$,
  where $z = N - k - j + 1$.

  For any $s < z$,

$$R_1(x - s\gamma_p T) - R_1(x + \gamma_m T - (s-1)\gamma_p T) = -T^2(\gamma_m + \gamma_p)$$

  and

$$V^{N-k+1}(x - z\gamma_p T) - V^{N-k+1}(x + \gamma_m T - (z-1)\gamma_p T) = -kT^2(\gamma_m + \gamma_p).$$

  Thus, it is optimal to switch ON the PBS if,

$$\begin{aligned} & R_1(x) - R_0(x) - (z-1)(\gamma_m + \gamma_p)T^2 - kT^2(\gamma_m + \gamma_p) && < && 0. \\ \iff \quad & CT - \gamma_m \frac{T^2}{2} - \gamma_p \frac{T^2}{2} - T^2(\gamma_m + \gamma_p)(z + k - 1) && < && 0 \ . \end{aligned}$$

$$\begin{aligned} \text{Maximum value of LHS } &= (k - z - k + 1)(\gamma_m + \gamma_p)T^2 \\ &= (k + j - N)(\gamma_m + \gamma_p)T^2, \\ &< 0 \quad (\because j \leq N - k) \end{aligned}$$

  Thus, the optimal policy is to switch OFF the PBS.

- $j + l \leq N - k$. There are two subcases here -

  - $l\gamma_p T < x < l\gamma_p T + x^*_{j+l}$

    The optimal policy would be to switch ON the PBS if -
    $R_1(x) + R_1(x - \gamma_p T) + \cdots + R_1(x - (l-1)\gamma_p T) + R_0(x - l\gamma_p T) + V^{j+l+1}(x + \gamma_m T - l\gamma_p T) < R_0(x) + R_1(x + \gamma_m T) + R_1(x + \gamma_m T - \gamma_p T) + \cdots +$
    $R_1(x + \gamma_m T - (l-2)\gamma_p T) + R_1(x + \gamma_m T - (l-1)\gamma_p T) + V^{j+l+1}(x + \gamma_m T - l\gamma_p T)$,
    which is true.

– $l\gamma_p T + x^*_{j+l} < x < (l+1)\gamma_p T$

The optimal policy would be to switch ON the PBS if -

$R_1(x) + R_1(x - \gamma_p T) + \cdots + R_1(x - (l-1)\gamma_p T) + R_1(x - l\gamma_p T) +$
$V^{j+l+1}(0) < R_0(x) + R_1(x + \gamma_m T) + R_1(x + \gamma_m T - \gamma_p T) + \cdots +$
$R_1(x + \gamma_m T - (l-1)\gamma_p T) + V^{j+l+1}(x + \gamma_m T - l\gamma_p T)$.

This is equivalent to

$$CT - (\gamma_m + \gamma_p)\frac{T^2}{2} - lT^2(\gamma_m + \gamma_p) + V^{j+l+1}(0) - V^{j+l+1}(x + \gamma_m T - l\gamma_p T) < 0.$$

The L.H.S is a decreasing function of x. Its value at $x = l\gamma_p T + x^*_{j+l}$ is
$CT - \frac{T^2}{2}(\gamma_m + \gamma_p)(2l+1) + V^{j+l+1}(0) - V^{j+l+1}(x^*_{j+l} + \gamma_m T)$.

From continuity of $V^{j+l}(x)$ at $x = x^*_{j+l}$, we have

$$R_1(x^*_{j+l}) + V^{j+l+1}(0) \leq R_0(x^*_{j+l}) + V^{j+l+1}(x^*_{j+l} + \gamma_m T).$$

(The inequality is applicable when $x^*_{j+l} = 0$.)

Thus, value of L.H.S at $\gamma_p T$

$$\begin{aligned} &\leq CT - \frac{T^2}{2}(\gamma_m + \gamma_p)(2l+1) + R_0(x^*_{j+l}) - R_1(x^*_{j+l}) \\ &= -lT^2(\gamma_m + \gamma_p), \\ &< 0. \end{aligned}$$

As the L.H.S is a decreasing function of x, its value is less than 0 throughout the range $l\gamma_p T + x^*_{j+l} < x < (l+1)\gamma_p T$. Thus, for all the above values of x, the optimal policy is to switch ON the PBS.

Thus, $a^*_j(x) = 1$ if $x > \gamma_p T$.

Combining all the above cases, we get $a^*_j(x) = Th_{x^*_j}(x)$ where $x^*_j < \gamma_p T$. □

Appendix

Appendix A - Review of Multimodularity

This section has been reproduced from [9] for a brief summary of concepts in multimodularity. More details can be found in [19]

Definition 1: Set $F := \{-e_1, s_2, s_3, \cdots, s_n, e_n\}$ *is the mulitmodular base where*

$$-e_1 = (\text{-1 } 0\ 0\ 0\ 0 \cdots\ 0\ 0), \;\; s_2 = (1 \text{ -1 } 0\ 0\ 0 \cdots\ 0\ 0)$$

$$s_3 = (0\ 1 \text{ -1 } 0\ 0 \cdots\ 0\ 0), \qquad s_4 = (0\ 0\ 1 \text{ -1 } 0 \cdots\ 0\ 0)$$

$$\vdots$$

$$s_N = (0\ 0\ 0\ 0\ 0 \cdots\ 1 \text{ -1}) \;\; and \;\; e_N = (0\ 0\ 0\ 0\ 0 \cdots\ 0\ 1).$$

Definition 2: A function $f : \{0,1\}^n \to R$ *is Multimodular if*

$$f_n(\mathbf{a}+\mathbf{v}) + f_n(\mathbf{a}+\mathbf{u}) \geq f_n(\mathbf{a}) + f_n(\mathbf{a}+\mathbf{u}+\mathbf{v})$$

for all $\mathbf{a} \in \{0,1\}^n$ *and for all* $\mathbf{u}, \mathbf{v} \in F$ *with* $\mathbf{u} \neq \mathbf{v}$ *and such that* $\mathbf{a}+\mathbf{u}$, $\mathbf{a}+\mathbf{v}$, $\mathbf{a}+\mathbf{u}+\mathbf{v} \in \{0,1\}^n$.

Definition 3: The bracket sequence $\boldsymbol{a}^*(\eta, \beta) := \{a_n(\eta, \beta)\}$ *with rate* $\eta \in [0,1)$ *and initial phase* $\beta \in [0,1)$ *is defined as*

$$a_n(\eta, \beta) = \lfloor n\eta + \beta \rfloor - \lfloor (n-1)\eta + \beta \rfloor.$$

Theorem 13. *A bracket sequence* $\boldsymbol{a}^*(\eta, \beta)$ *for any* $\beta \in [0,1)$ *minimizes the cost*

$$\underset{N\to\infty}{lim\ sup} \tfrac{1}{N} \textstyle\sum_{n=1}^{N} f_n(a_1, \cdots, a_n)$$

over all the sequences that satisfy

$$\underset{N\to\infty}{lim\ inf} \tfrac{1}{N} \textstyle\sum_{n=1}^{N} a_n \geq \eta$$

where $\eta \in [0, 1)$, under the following assumptions:

1) f_n is multimodular $\forall$ n.

2) $f_n(a_1, \cdots, a_n) \geq f_{n-1}(a_2, \cdots, a_n)$ $\forall$ $n > 1$

3) $\forall$ sequence$\{a_n\}$, $\exists$ a sequence$\{b_n\}$ $\forall$ n, m with $n > m$, such that

$$f_n(b_1, \cdots, b_{n-m}, a_1, \cdots, a_m) = f_m(a_1, \cdots, a_m)$$

4) $\forall$ n, the functions$f_n(a_1, \cdots, a_n)$ are increasing in $a_i \forall$ i. $\square$

Abbreviations

3GPP	:	3rd Generation Partnership Project
BS	:	Base Station
IID	:	Independent and Identical Distributions
MBS	:	Macro Base station
MDP	:	Markov Decision Process
MMAP	:	Markov Modulated Arrival Process
PBS	:	Pico Base station
QoS	:	Quality of service

Bibliography

[1] International Telecommunication Union, “The World in 2014: ICT Facts and Figures,” Tech. Rep., 2014. [Online]. Available: http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2014-e.pdf

[2] eMarketer, “Worldwide Mobile Phone Users: H1 2014 Forecast and Comparative Estimates ,” eMarketer, Tech. Rep., 2014.

[3] Third Generation Partnership Project, “3GPP TR 36.814 V9.0.0 (Release 9),” Tech. Rep., Mar 2010. [Online]. Available: http://www.qtc.jp/3GPP/Specs/36814-900.pdf

[4] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, “Green Cellular Networks: A Survey, Some Research Issues and Challenges,” *IEEE Communications Surveys and Tutorials*, vol. 30, no. 4, pp. 524–540, 2011.

[5] L. Saker, S. E. Elayoubi, R. Combes, and T. Chahed, “Optimal Control of Wake Up Mechanisms of Femtocells in Heterogeneous Networks,” *IEEE Journal on Selected Areas in Communication*, vol. 30, no. 3, pp. 664–672, 2012.

[6] Z. Niu, Y. Wu, J. Gong, and Z. Yang, “Cell Zooming for Cost-Efficient Green Cellular Networks,” *IEEE Communications Magazine*, vol. 48, no. 11, pp. 74–79, 2010.

[7] G. Micallef, P. Mogensen, and H. Scheck, “Cell Size Breathing and Possibilities to Introduce Cell Sleep Mode ,” *European Wireless Conference (EW)*, pp. 111–115, 2010.

[8] I. Ashraf, F. Boccardi, and L. Ho, “Power Savings in Small Cell Deployments via Sleep Mode Techniques,” *21st International Symposium on Personal, Indoor and Mobile Radio Communications Workshops (PIMRC Workshops)*, pp. 307–311, 2010.

[9] S. Ramnath, V. Kavitha, and E. Altman, “Open Loop Optimal Control of Base Station Activation for Green Networks,” *International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pp. 161–166, 2011.

[10] P. Chunyi, L. Suk-Bok, L. Songwu, L. Haiyun, and L. Hewu, “Traffic-Driven Power Saving in Operational 3G Cellular Networks,” *17th Annual International Conference on Mobile Computing and Networking (MobiCom)*, pp. 121–132, 2011.

[11] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, “Towards Dynamic Energy Efficient Operation of Cellular Network Infrastructure,” *IEEE Communications Magazine*, vol. 49, no. 6, pp. 56–61, 2011.

[12] L. Saker, S. E. Elayoubi, and H. O. Scheck, "System Selection and Sleep Mode for Energy Saving in Cooperative 2G/3G Networks," *70th Vehicular Technology Conference Fall (VTC 2009-Fall)*, pp. 1–5, 2009.

[13] S. E. Elayoubi, L. Saker, and T. Chahed, "Optimal Control for Base Station Sleep Mode in Energy Efficient Radio access Networks," *IEEE INFOCOM*, pp. 106–110, 2011.

[14] V. Kavitha, S. Ramnath, and E. Altman, "Spatial Queueing for Analysis, Design and Dimensioning of Picocell Networks with Mobile Users," *Elseiver Performance Evaluation*, vol. 68, no. 8, pp. 710–727, 2011.

[15] V. Kavitha, S. Ramnath, and E. Altman, "Analysis of Small Cell Networks with Randomly Wandering Users ," *International Symposium on Modeling and Optimization of Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pp. 60–67, 2012.

[16] R. W. Wolff, *Stochastic Modelling and the Theory of Queues.* Prentice Hall, 1989.

[17] B. Hajek, "Extremal Splittings of Point Processes," *Mathematics of Operations Research*, vol. 10, no. 4, pp. 543–556, 1985.

[18] M. Agarwal, V. Borkar, and A. Karandikar, "Structural Properties of Optimal Transmission Policies over a Randomly Varying Channel," *IEEE Transactions on Automatic Control*, vol. 53, no. 6, pp. 1476 – 1491, 2008.

[19] E. Altman, B. Goujal, and A. Hordijk, *Discrete-Event Control of Stochastic Networks Multimodularity and Regularity.* Springer, 2001.

[20] Global e-Sustainability Initiative, "GeSI SMART 2020 Report," Tech. Rep., 2008. [Online]. Available: http://gesi.org/SMARTer2020

[21] International Telecommunication Union, "The World in 2009: ICT Facts and Figures," Tech. Rep., 2009. [Online]. Available: http://www.itu.int/ITU-D/ict/material/Telecom09-flyer.pdf

Index